MULTIPLE CRITERIA DECISION MAKING

ANNUAL

VOL. 12 (2017)

THE UNIVERSITY OF ECONOMICS IN KATOWICE 2017

Printed by

Drukarnia Archidiecezjalna w Katowicach, ul. Wita Stwosza 11, 40-042 Katowice
www.drukarch.com.pl

ISSN 2084-1531

Edition: 100 copies

Original version of the MCDM is the paper version

Publisher of The University of Economics in Katowice
ul. 1 Maja 50, 40-287 Katowice, tel. +48 32 25 77 635, fax +48 32 25 77 643
www.wydawnictwo.ue.katowice.pl, e-mail: wydawnictwo@ue.katowice.pl

Contents

From the Editor

International Workshops on Multiple Criteria Decision Making are scientific conferences organized every two years by the Department of Operations Research, University of Economics in Katowice, Poland. They are devoted to the theory and applications of multiobjective optimization, goal programming and multiple criteria decision aid. Papers presented at the workshops discuss various practical problems solved by MCDM methods in economics as well as in construction ecology, transportation, health care, education and other fields. After a presentation at the workshop and the blind review process, they are often published in the journal *Multiple Criteria Decision Making*. Selected papers, presented at the Seventh International Workshop on Multiple Criteria Decision Making held in Ustroń, Poland, on 2-4 April, 2017, are included in the first part of this volume. The second part contains regularly contributed papers.

In the paper ***Bicriteria optimization in the risk-adjusted newsvendor problem*** (**M. Bieniek**), a problem in which the first objective is the classical maximization of the expected profit and the second one is the satisficing-level objective is studied with various degrees of risk tolerance. The formulas obtained are illustrated with exponentially distributed demand.

The paper ***Multicriteria analysis of the success of research projects*** (**B. Gładysz**, **D. Kuchta**) shows that Data Envelopment Analysis is an important tool for the evaluation of R&D activities.

In the paper ***BIPOLAR MIX – a method for mixed evaluations and its application to the ranking of European projects***, **D. Górecka** presents a new discrete MCDA tool developed for mixed performance of alternatives and its application.

The paper ***Trade-off guided search for approximate Pareto optimal portfolios*** (**P. Juszczuk, I. Kaliszewski, J. Miroforidis**) attempts to represent the Pareto Front in the Markowitz mean-variance model by two-sided discrete approximations.

The paper ***Supporting multicriteria fuzzy decisions on the Forex market*** (**P. Juszczuk, L. Kruś**) deals with decisions made by a decision maker using technical analysis indicators. The method proposed allows to extend the number of currency pairs analyzed, without significantly increasing the computation time.

In the paper ***Comparing the crisp and fuzzy approaches to modelling preferences towards health states***, **B. Kamiński and M. Jakubczyk** test whether treating the impact of health worsening (defined using the EQ-5D-5L descriptive system, i.e. decomposing health status into five criteria) as a fuzzy concept can improve the model fit.

In the paper ***New results on the quality of recently introduced index for a consistency control of pairwise judgments***, **P.T. Kazibudzki** examines the efficiency of a recently proposed consistency index based on the redefined idea of triads inconsistency within Pairwise Comparison Matrices. The quality of the proposal is studied and compared to other ideas with the application of Monte Carlo simulations.

In the paper ***Multicriteria assessment of the academic research activity*** (**D. Kuchta, R. Ryńca, Y. Ziaeian, A. Grudziński**) a network DEA approach to deal with efficiency assessment is presented and applied to the assessment of performance of members of an academic faculty of Wroclaw University of Science and Technology.

In the paper ***Assessing the strategic factors and choosing the development scenarios for local administrative units using AHP***, **A. Łuczak** studies strategic factors (objectives, tasks and development scenarios) and selects the best scenario for local administrative units.

In the paper ***The use of initial filters to direct search in decision processes***, **D.M. Ramsey** considers strategies based on the filtering of initial information. A new model is presented according to which the goal of the decision maker is to maximize her or his expected reward from search taking into account the search costs.

The paper ***DEMATEL as a weighting method in multi-criteria decision analysis*** (**A. Kobryń**) is the first of the regularly contributed papers. A new weighting procedure is proposed and verified.

The paper ***Interactive procedure for multiobjective dynamic programming with the mixed ordered structure*** (**M. Nowak, S. Sitarz, T. Trzaskalik**) is the second of the regularly contributed papers. It presents a multiobjective dynamic programming problem with the values of the criteria function in ordered structures. The ordered structures and the proposed procedure are illustrated by numerical examples.

Tadeusz Trzaskalik

Part I

Papers presented during the Seventh International Workshop on Multiple Criteria Decision Making

MULTIPLE CRITERIA DECISION MAKING

Vol. 12 2017

Milena Bieniek*

BICRITERIA OPTIMIZATION IN THE RISK-ADJUSTED NEWSVENDOR PROBLEM

DOI: 10.22367/mcdm.2017.12.01

Abstract

In this paper, we study the newsvendor problem with various degrees of risk tolerance. We consider bicriteria optimization where the first objective is the classical maximization of the expected profit and the second one is the satisficing-level objective. The results depend on the risk coefficient and are different for a risk-neutral, a risk-averse, and a risk-seeking retailer. We find the compromise solution of the bicriteria newsvendor problem numerically, since the two objectives are mutually conflicting. The formulas obtained are illustrated with exponentially distributed demand.

Keywords: stochastic demand, newsvendor problem, bicriteria optimization, risk.

1 Introduction

The single-period newsvendor problem is one of the most fundamental inventory models (cf. Silver et al., 1998). In the classical version of this problem the objective is to maximize the expected profit, but many other objectives can be used. A survey of this topic has been performed recently by Qin et al. (2011). Sometimes, it is more relevant to consider the retailer's risk tolerance. In the newsvendor model, various degrees of risk can be assumed. The most popular measures assuming risk aversion are Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). Lately, Teunter et al. (2013a) have studied how the capacity for uncertainty influences inventory decisions of a risk-averse newsvendor using the VaR and CVaR criteria. These criteria have also been studied by Teunter et al.

* Maria Curie-Skłodowska University, Economic Faculty, Department of Statistics and Econometrics, Lublin, Poland, e-mail: milena.bieniek@poczta.umcs.lublin.pl.

(2013b), but with uncertainty in the shortage cost. Moreover, in Teunter et al. (2014), under the CVaR criterion, the authors obtain the optimal quantity and pricing decisions under both quantity and pricing competition. The most recent papers in this field are, among others, Őzler (2009), Wang and Webster (2009), Xinshenga et al. (2015), Rubio-Herrero et al. (2015), Dia and Mengb (2015), Ray and Jenamani (2016) and Ye and Sun (2016).

The degree of risk tolerance is studied in Arcelus et al. (2012b) and Raza et al. (2017), too. In these papers, the authors consider the newsvendor model with random and price-dependent demand. Among other objectives, they use the risk coefficient into the satisficing-level objective. In general, the satisficing-level objective is defined as the maximization of the probability of the event that the profit is greater than or equal to the prespecified target profit. The satisficing-level objective in the newsvendor problem is treated, for instance, in Kabak and Shiff (1978), Lau (1980), Li et al. (1991), Yang (2011) or Pinto (2016). The satisficing-level objective with a moving target and price independent demand is explored in Parlar and Weng (2003), Arcelus et al. (2012a) and Bieniek (2016), but for a risk-neutral retailer. The moving target considered in these papers is the expected profit.

Here we continue the study of a similar problem, but with a risk-adjusted retailer. It should be noted that the risk-adjusted expected profit is defined in Arcelus et al. (2012b). Moreover, in this paper, the probability of the event that the classical profit is greater than or equal to the risk-adjusted expected profit is maximized for uniformly distributed demand. We analyse a more relevant and more generalized objective, where the profit is replaced by the risk-adjusted profit (a notion introduced in our paper). This is a more appropriate approach to the matter since we study the preferences of the retailer of each kind. Additionally, we solve the satisficing-level risk-adjusted newsvendor problem for general distribution. As a result, we obtain approximate solutions which are strictly dependent on the risk coefficient. We also apply the exponential distribution, which is widely used in practice, to the results obtained. Since the exponential distribution is mathematically tractable, we are able to obtain exact solutions to the problem. It should be emphasized here that the use of the uniform distribution in Arcelus et al. (2012b) also gives exact results, but in real life there are no products whose demand can be modelled by this distribution. Moreover, in our opinion, in the paper cited the solution is not complete because some special cases of the problem should be added and the solution should depend on the risk coefficient. This gap can be complemented by our paper.

Furthermore, in our study we combine the satisficing-level objective with the classical expected profit objective into the bicriteria index. Here the classical objective is to maximize the risk-adjusted expected profit and the satisficing-level objective is to maximize the probability that the risk-adjusted profit is greater than or equal to the risk-adjusted expected profit. Since these two objectives are mutually conflicting, we find the compromise solution which can be done numerically.

The rest of the paper is organized as follows. Section 2 is devoted to the general bicriteria risk-adjusted newsvendor problem and provides the basic notation and definitions. The notion of a bicriteria index is also recalled. The risk coefficient is recalled and the notion of the risk-adjusted profit is introduced. Approximate solutions of the satisficing-level model and the bicriteria model are presented for the general distribution. In Section 3, the exponential distribution is applied to the results obtained, which allow us to give exact solutions. Next, we illustrate the formulas obtained by a numerical example and draw graphs of auxiliary functions. Finally, we perform sensitivity analysis of the changes of the risk coefficient.

2 The bicriteria newsvendor problem with the risk-adjusted profit – general case

In this section, we recall the definitions of the profit function and the risk-adjusted expected profit, and introduce the definition of the risk-adjusted profit. Using these quantities, we solve the newsvendor problem with a classical risk-adjusted expected profit objective and with a risk-adjusted satisficing-level objective. Finally, we investigate the bicriteria problem with both these objectives.

In the classical newsvendor problem, we examine a retailer who wants to acquire Q units of a given product subject to random demand. We use the following notation. Let:

- $p > 0$ be the selling price for unit (unit revenue);
- $c > 0$ be the purchasing cost per unit;
- $s > 0$ be the unit shortage cost;
- v be the unit salvage value (unit price of disposing any excess inventory);
- $f(.)$ and $F(.)$ be the probability density function and the cumulative distribution function of the demand with mean μ;
- $\lambda \geq 0$ be the risk coefficient.

The standard assumption is $v < c < p$. The risk coefficient expresses the risk tolerance of the retailer. There are four risk categories. Namely, for $\lambda = 0$ we have a riskless retailer and for $\lambda = 1$, a risk-neutral retailer. For $0 < \lambda < 1$ we are dealing with a risk-seeker and for $\lambda > 1$, with a risk-averse retailer (cf. Arcelus et al., 2012b).

We define the risk-adjusted profit by the formula:

$$\pi_\lambda(Q, x) = \begin{cases} (p - c)x - \lambda(c - v)(Q - x), if\ x \leq Q \\ (p - c)x - \lambda(p + s - c)(x - Q), if\ x > Q, \end{cases}$$

where Q is the order quantity and x is the realized demand. Then the risk-adjusted expected profit E_λ is equal to:

$$E_\lambda(Q) = (p-c)\mu - \lambda\left[(c-v)(Q-\mu) + (p+s-v)\int_Q^\infty (x-Q)f(x)dx\right]. \quad (1)$$

Arcelus et al. (2012b) justify using the risk coefficient as follows. They state that the first term in the formula for the risk-adjusted expected profit, without the risk coefficient, stands for certain gains. The second term, with the risk coefficient, indicates uncertain losses and includes the variability of the random demand. From this definition, we can further see that the higher the degree of risk-aversion, the higher the value of the risk coefficient.

Now, if the objective is to maximize the risk-adjusted expected profit, then this model gives the same optimal order quantity as the model for a risk-neural retailer and does not depend on λ. Because of that the order quantity maximizing the risk-adjusted expected profit Q_E^* can be obtained from:

$$F(Q_E^*) = (p+s-c)/(p++s-v).$$

But, in the satisficing-level model, where the objective is to maximize the survival probability, the results depend on the risk coefficient. Here the so-called survival probability $H_\lambda(Q)$ is the probability of the event that the risk-adjusted profit is greater than or equal to the risk-adjusted expected profit, namely:

$$H_\lambda(Q) = P\big(\pi_\lambda(Q) \geq E_\lambda(Q)\big).$$

Let Q_H^* be the optimal order quantity which maximizes $H_\lambda(Q)$. The following theorem is crucial for the subsequent analysis because it gives the possible expressions for the survival probability.

Theorem 1

1. If $[(p-c)(1-\lambda) - \lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$ then the profit function $\pi_\lambda(Q,x)$ is increasing-decreasing as a function of the realized demand, and the survival probability is equal to:
$$H_\lambda(Q) = \int_{D_1(Q)}^{D_2(Q)} f(x)dx.$$
2. If $(p-c)(1-\lambda) - \lambda s > 0$ and $\lambda < 1$ then the profit function $\pi_\lambda(Q,x)$ is increasing as a function of the realized demand, and the survival probability:
   a) for $E_\lambda(Q) < (p-c)Q$ is given by:
$$H_\lambda(Q) = \int_{D_1(Q)}^{\infty} f(x)dx,$$
   b) for $E_\lambda(Q) \geq (p-c)Q$ is given by:
$$H_\lambda(Q) = \int_{D_2(Q)}^{\infty} f(x)dx,$$

where $D_1(Q) = \max\{0, k(Q)\}$ with:

$$k(Q) = \frac{(c-v)Q + E_\lambda(Q)}{p-v}$$

and $D_2(Q) = \max\{0, l(Q)\}$ with:

$$l(Q) = \frac{(p+s-c)Q - E_\lambda(Q)}{s}.$$

Proof

In this theorem, as compared with the earlier results by Parlar and Weng (2003), the new case 2) occurs. This is a consequence of introducing the risk coefficient. Thus, it is enough to prove this case. Note that if $(p-c)(1-\lambda) - -\lambda s > 0$ and $\lambda < 1$ then the slope of the profit function is positive and $E_\lambda(Q)$ can be greater than $(p-c)Q$, which proves the desired result.

In Figure 1 we illustrate this theorem by the graphs of the profit function as a function of the realized demand.

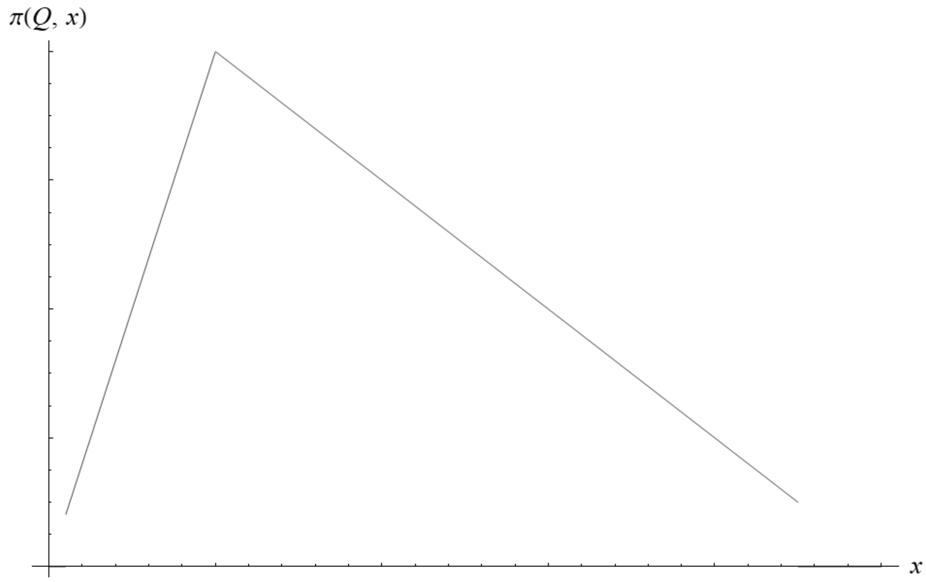


Figure 1. Profit function $\pi_\lambda(Q, x)$ for $[(p-c)(1-\lambda) - \lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$

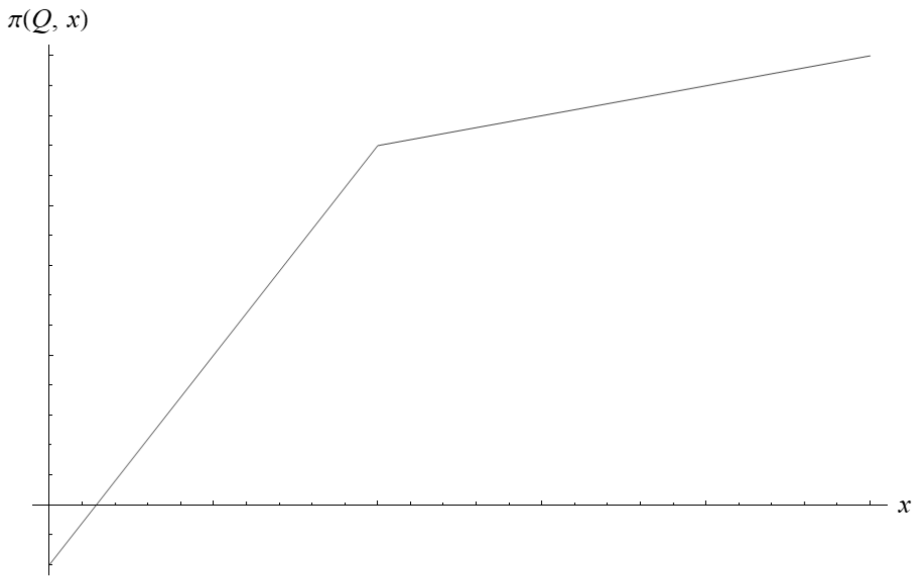


Figure 2. Profit function $\pi_\lambda(Q, x)$ for $(p - c)(1 - \lambda) - \lambda s > 0$ and $\lambda < 1$

Remark 1

Using formula (1), the expressions in the theorem can be written as:

$$k(Q) = \mu - \frac{\lambda(p + s - v)\int_Q^\infty (x - Q)f(x)dx}{p - c + \lambda(c - v)}$$

and:

$$l(Q) = \frac{(p - c + \lambda(c - v))\mu - \lambda(p + s - v)[Q + \int_Q^\infty (x - Q)f(x)dx]}{p - c - \lambda(p + s - c)}.$$

Remark 2

Case b) of Theorem 1 is satisfied for a risk-seeker and for products for which $(p - c)(1 - \lambda)/\lambda > s$. It corresponds to the situation when the shortage cost per item is smaller than the maximal profit per item multiplied by $\frac{1-\lambda}{\lambda}$.

Now, we examine the variability of the survival probability. To this aim it is necessary to analyse the variability of the limit functions. First, we explore the monotonicity and the zeros of these functions. Note that all results involve the risk coefficient.

For the lower limit, we have:

$$k(0) = \frac{(p - c)(1 - \lambda) - \lambda s}{p - c + \lambda(c - v)}\mu.$$

Thus $k(0) > 0$ for $[(p - c)(1 - \lambda) - \lambda s > 0$ and $\lambda < 1]$ and $k(0) < 0$ for $[(p - c)(1 - \lambda) - \lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$. Moreover:

$$k'(Q) = \frac{\lambda(p+s-v)}{p-c+\lambda(c-v)}(1-F(Q)) > 0$$

and:

$$k''(Q) = -\frac{\lambda(p+s-v)}{p-c+\lambda(c-v)}f(Q) < 0,$$

which implies that $k(Q)$ is concave and increasing. For $[(p-c)(1-\lambda)-\lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$ let Q_{01} be such that $k(Q_{01}) = 0$. This implies that $D_1(Q)$ is equal to 0 on the interval $(0, Q_{01})$ and is an increasing function of the order quantity for $Q \geq Q_{01}$. Moreover, it tends to μ as Q tends to infinity.

For the upper limit, we have:

$$l(0) = \mu,$$

$$l'(Q) = -\frac{\lambda(p+s-v)}{(p-c)(1-\lambda)-\lambda s}F(Q),$$

$$l''(Q) = -\frac{\lambda(p+s-v)}{(p-c)(1-\lambda)-\lambda s}f(Q).$$

Let Q_{02} be such that $l(Q_{02}) = 0$. If $(p-c)(1-\lambda)-\lambda s > 0$ and $\lambda < 1$ then the upper limit $D_2(Q)$ is decreasing for $Q \leq Q_{02}$ and equal to 0 for $Q > Q_{02}$, otherwise $D_2(Q)$ is increasing and tends to infinity as Q tends to infinity.

Additionally, for $[(p-c)(1-\lambda)-\lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$ if:

$$D_2'(Q_M) - D_1'(Q_M) = 0,$$

then we infer that the minimum distance between the limit functions is attained for Q_M such that $F(Q_M) = [(p-c)(\lambda-1)+\lambda s]/[\lambda(p+s-v)]$ (cf. Parlar and Weng, 2003).

Using the above properties, we obtain the approximate solution for the satisficing-level model which is later used in the bicriteria model. The next theorem holds for a risk-adjusted retailer. It is equivalent to the theorem for a risk-neutral retailer presented in Parlar and Weng (2003). In our case, the solution involves the risk coefficient and one more additional case occurs.

Theorem 2

1. If $[(p-c)(1-\lambda)-\lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$ and if for some parameters p, s, v we have:

$$a(Q) < b(Q) \; for \; Q > Q_M,$$

where $a(Q) = f(D_2(Q))/f(D_1(Q))$ and $b(Q) = \frac{\lambda s-(1-\lambda)(p-c)}{p-\lambda v-(1-\lambda)c}\frac{1-F(Q)}{F(Q)}$, then the survival probability $H_\lambda(Q)$ is decreasing on (Q_M, ∞) and attains its maximum on the interval (Q_{01}, Q_M).
2. If $(p-c)(1-\lambda)-\lambda s > 0$ and $\lambda < 1$ then the survival probability $H_\lambda(Q)$ attains its maximum at Q_1, defined by $E_\lambda(Q_1)-(p-c)Q_1 = 0$. Then the maximal survival probability is equal to $1-F(Q_1)$.

Proof

It is enough to prove 2). In this case, the risk-adjusted profit function is an increasing function of the realized demand. Furthermore, $E_\lambda(0) = [(p-c)(1-\lambda) - \lambda s\mu > 0$ and, additionally, $E\lambda(Q)$ is a convex function for all $Q>0$. This implies that there exists Q_1 such that $E_\lambda > (p-c)Q$ for $Q < Q_1$, $E_\lambda < (p-c)Q$ for $Q > Q_1$ and $E_\lambda(Q_1) = (p-c)Q_1$. These inequalities are illustrated in Figure 3. Therefore, we have $D_1(Q_1) = D_2(Q_1) = Q_1$. Since $D_2(Q)$ is in this case a decreasing function of Q and $D_1(Q)$, an increasing function of Q, then using Theorem 1b we see that the optimal order quantity is Q_1 and the maximum survival probability is $H_\lambda(Q_1) = \int_{Q_1}^{\infty} f(x)dx$, which gives the desired result.

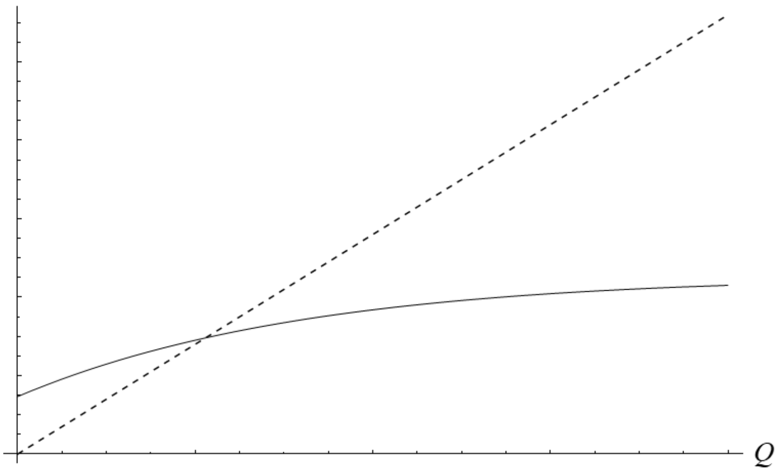


Figure 3. Functions: $(p-c)Q$ (dashed) and $E_\lambda(Q)$ (solid) for $(p-c)(1-\lambda) - \lambda s > 0$ and $\lambda < 1$

Now we combine both objectives in the bicriteria newsvendor model. These objectives are mutually conflicting. For this reason, the bicriteria index $Y(Q)$ is used, which is defined by:

$$Y_\lambda(Q) = \frac{w}{E_\lambda^*} E_\lambda(Q) + \frac{1-w}{H_\lambda^*} H_\lambda(Q)$$

(cf. Parlar and Weng, 2003). Let $E_\lambda^* = E_\lambda(Q_E^*)$ and $H_\lambda^* = H_\lambda(Q_H^*)$. Note also that both E_λ^* and H_λ^* are constants in the bicriteria function. The weight $0 \leq w \leq 1$ measures the relative importance of $E_\lambda(Q)$ and $H_\lambda(Q)$. Our aim is to find the order quantity which maximizes the bicriteria index. It can be considered as a compromise solution to the bicriteria problem. Let us recall that $Q_H^* < Q_E^*$ holds always and since the function $Y_\lambda(Q)$ is continuous, it attains its maximum on the interval (Q_H^*, Q_E^*). The derivative of the bicriteria index is equal to

$Y'(Q) = \frac{w}{E_\lambda^*} E_\lambda'(Q) + \frac{(1-w)}{H_\lambda^*} H_\lambda'(Q)$ and it suffices to find Q_Y^* such that $Y'(Q_Y^*) = 0$ and then to prove that $Y''(Q) < 0$ for all $Q > Q_H^*$. As a result, we obtain a unique Q_Y^* dependent on λ which maximizes the bicriteria index and satisfies the inequality $Q_H^* \leq Q_Y^* \leq Q_E^*$. From now on we will write $Y^* = Y(Q_Y^*)$. Note that if the second derivative $H''(Q) > 0$ then $Y''(Q) < 0$for the weight w such that:

$$w > \frac{E_\lambda^* H_\lambda''(Q)}{E_\lambda^* H_\lambda''(Q) - H_\lambda^* E_\lambda''(Q)},$$

where all $Q > Q_H^*$. The compromise solution optimal for the bicriteria problem can be found numerically.

3 The newsvendor problem with the risk-adjusted profit – exponential demand case

This section shows the results of the previous section for exponentially distributed demand. Exact solutions and numerical examples are given for demand with the density $f(x) = \alpha e^{-\alpha x},\ x > 0$ and the cumulative distribution function $F(x) = 1 - e^{-\alpha x}$, $x > 0$, where $\alpha > 0$ is the parameter of this distribution. The mean demand is $\mu = \frac{1}{\alpha}$. In this case, the order quantity maximizing the risk-adjusted expected profit is $Q_E^* = \frac{1}{\alpha} \ln \frac{p+s-v}{c-v}$.

Since for the exponential distribution we have $Q_H^* = Q_{01} = Q_M$, which was proved in Bieniek (2016), the counterpart of the Theorem 2 is the following:

Theorem 3
If the demand distribution in the newsvendor problem is an exponential distribution with the parameter $\alpha > 0$, then the following statements hold.

1. If $[(p-c)(1-\lambda) - \lambda s < 0$ and $\lambda < 1]$ or $\lambda > 1$ and if for some parameters p, s, v we have:

$$a(Q) < b(Q)\ \ for\ Q > Q_0,$$

where $a(Q) = e^{-\lambda(D_2(Q) - D_1(Q))}$ and $b(Q) = \frac{\lambda s - (1-\lambda)(p-c)}{p - \lambda v - (1-\lambda)c} \frac{e^{-\alpha Q}}{1 - e^{-\alpha Q}}$, then then the survival probability attains its maximum value at $Q_H^* = \frac{1}{\alpha} \ln \frac{p+s-v}{c-v+(p-c)/\lambda}$ and this maximum value is equal to:

$$H^* = H(Q_H^*) = 1 - \left(\frac{\frac{p-c}{\lambda} + c - v}{p - v + s} \right)^{\frac{\lambda(p-v+s)}{(p-c)(\lambda-1)+\lambda s}}.$$

2. If $(p-c)(1-\lambda) - \lambda s > 0$ and $\lambda < 1$ then the survival probability attains its maximum at Q_1, such that $E_\lambda(Q_1) - (p-c)Q_1 = 0$. Then $H^* = e^{-\alpha Q_1}$.

Now we illustrate the results of Theorem 3 with a graph. Figure 4 shows the graph of of $H(Q)$ for the parameters $(\alpha, v, c, p, s) = (0.003, 15, 16, 30, 50)$. We assume that the risk coefficient is equal to 1.3, 1.0 and 0.7, respectively.

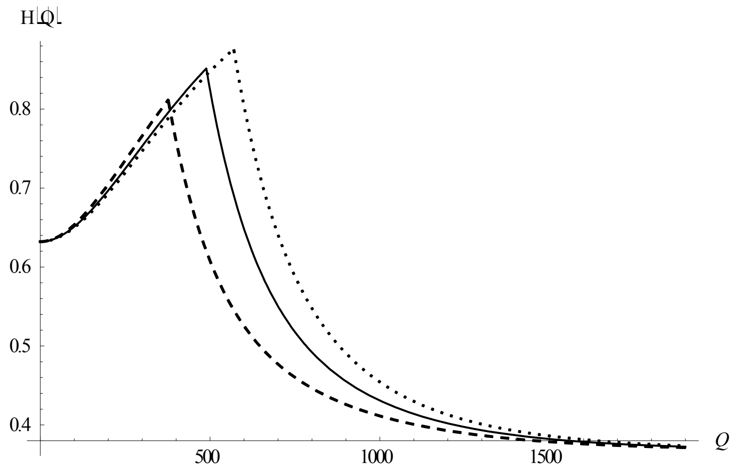


Figure 4. Survival probability H(Q) for $(\alpha, v, c, p, s) = (0.003,15,16,30,50)$) with $\lambda = 1.3$ (dotted); $\lambda = 1.0$ (solid) and $\lambda = 0.7$ (dashed)

We conclude that for a more risk-averse retailer the order quantity maximizing the survival probability increases and the maximal survival probability also increases. Furthermore, for a more risk-seeking newsvendor the optimal order quantity decreases along with the optimal survival probability.

In what follows, we present a numerical example with the same parameter values. The optimal solution of the classical newsvendor model and the satisficing-level model are calculated separately, for various values of λ. The risk coefficient λ is fixed and varies from 0.2 to 1.3 with step 0.1.

Table 1: Sensitivity analysis for various parameters λ for $(\alpha, v, c, p, s) = (0.003,15,16,30,50)$

λ	Q_H^*	H^*	Q_E^*	E^*
0.7	376.622	0.8116	1391.46	3692.64
0.8	418.872	0.8274	1391.46	3553.5
0.9	455.889	0.8404	1391.46	3414.35
1.0	488.779	0.8514	1391.46	3275.2
1.1	518.334	0.8607	1391.46	3136.06
1.2	545.138	0.8688	1391.46	2996.91

Let us analyse the results presented in Table 1. We see that the optimal order quantity maximizing the risk adjusted expected profit is always larger (possibly even three-fold) than the solution of the satisficing-level model. If the risk coefficient increases from 0.7 to 1.2 then the optimal order quantity maximizing the survival quantity increases from 376.622 to 545.138. Moreover, the optimal order quantity in the classical objective does not depend on the risk coefficient and remains constant, equal to 1391.46 for any λ. But both the maximum survival probability and the risk adjusted expected profit depend on the risk coefficient. The survival probability increases from 0.811 to 0.8688 if λ increases from 0.7 to 1.2. By contrast, the maximum risk-adjusted expected profit decreases from 3692.64 to 2996.91 for the increasing risk coefficient.

Now we show the results of a numerical example of the bicriteria problem. Note that the optimal order quantity Q_Y^* can be found numerically. We assume that the weight increases from 0 to 1 with step 0.1. The risk coefficient λ is set to 0.8, 1.0 and 1.2, respectively. We examine the sensitivity of the optimal solution with respect to weight w. The model parameters here are also $(\alpha, v, c, p, s) =$ $= (0.003, 15, 16, 30, 50)$. For these parameters to ensure the negativity of $Y''(Q)$, weight w should be greater than or equal to 0.5. For $w < 0.5$ we simply take $Q_Y^* = Q_H^*$.

Table 2: Sensitivity analysis for various values of weight w
for $(\alpha, v, c, p, s) = (0.003,15,16,30,50)$

$\boldsymbol{\lambda}$	**0.8**	**0.8**	**1.0**	**1.0**	**1.2**	**1.2**
$\boldsymbol{w}$	$\boldsymbol{Q_Y^*}$	$\boldsymbol{Y^*}$	$\boldsymbol{Q_Y^*}$	$\boldsymbol{Y^*}$	$\boldsymbol{Q_Y^*}$	$\boldsymbol{Y^*}$
0.5	1300.06	0.7334	1310.09	0.7289	1317.43	0.7266
0.6	1333.06	0.7863	1339.52	0.7827	1344.2	0.7808
0.7	1354.96	0.8396	1359.01	0.8368	1361.93	0.8354
0.8	1370.6	0.893	1372.91	0.8911	1374.58	0.8902
0.9	1382.33	0.9464	1383.34	0.9455	1384.07	0.945
1.0	$Q_E^* = 1391.46$	1.0	Q_E^*	1.0	Q_E^*	1.0

We note that as weight w increases, the optimal order quantity maximizing the bicriteria index also increases. Moreover, for greater values of w the expected profit model has an increasing influence on the bicriteria model. For this reason, the optimal value Q_Y^* is closer to the optimal order quantity Q_E^* of the expected profit model. Otherwise, lower values of weight imply a greater influence of the satisficing-level model on the bicriteria model. Additionally, for $w < 0.5$ the order quantity maximizing the bicriteria index is assumed to be equal to the optimal order quantity maximizing the survival probability. All these statements hold for a certain fixed value of the risk coefficient.

4 Conclusions

Our paper deals with the bicriteria optimization in the newsvendor problem with various risk tolerance. First, we define the risk-adjusted profit function. The first objective investigated is the classical objective of the maximization of the expected profit but for the risk-adjusted profit. The second objective considered is the maximization of the probability that the risk-adjusted profit is greater than or equal to the risk-adjusted expected profit. We determine the solutions separately for each model and then we obtain the compromise solution with two conflicting objectives. The solution of the bicriteria problem can be found numerically. Since the results of the satisficing-level model in the general case are approximated, we use the exponential distribution to determine exact solutions. All the results strictly depend on the risk tolerance of the retailer. Namely, the solutions are different for a risk-averse, a risk-neutral, and a risk-seeking newsvendor. The sensitivity analysis of the changes of the risk coefficient is also performed.

Future research can include, for instance, the creation of new algorithms for the solution of this model. Other modifications of the profit target setting, taking into account the behaviour of the customer, could also bring interesting results.

References

Arcelus F.J., Kumar S., Srinivasan G. (2012a), *Bicriteria Optimization in the Newsvendor Problem with Uniformly Distributed Demand*, 4OR Journal of Operations Research, 10, 267-286.

Arcelus F.J., Kumar S., Srinivasan G. (2012b), *Risk Tolerance and a Retailer's Pricing and Ordering Policies within a Newsvendor Framework*, Omega, 40, 188-198.

Bieniek M. (2016), *Bicriteria Optimization in the Newsvendor Problem with Exponentially Distributed Demand*, Multiple Criteria Decision Making, 11, 20-35.

Choi T.-M., ed. (2012), *Handbook of Newsvendor Problems*, International Series in Operations Research and Management Science, Vol. 176, Springer, New York, Heidelberg, Dordrecht.

Dia J., Mengb W. (2015), *A Risk-averse Newsvendor Model under Marketing-dependency and Price-dependency*, International Journal of Production Economics, 160, 220-229.

Gallego G., Katricoglu K., Ramachandran B. (2007), *Inventory Management under Highly Uncertain Demand*, Operations Research Letters, 35, 281-289.

Kabak I., Shiff A. (1978), *Inventory Models and Management Objectives*, Sloan Management Review, 19(2), 53-59.

Lau H.-S. (1980), *The Newsboy Problem under Alternative Optimization Objectives*, Journal of Operations Research Society, 31, 525-535.

Li J., Lau H., Lau A. (1991), *A Two-product Newsboy Problem with Satisficing Objective and Independent Exponential Demands*, IIE Transactions, 23, 29-39.

Őzler A., Tan A., Karaesmen F. (2009), *Multi-product Newsvendor Problem with Value-at-Risk Considerations*, International Journal of Production Economics, 117(2), 244-255.

Parlar M., Weng Z.K. (2003), *Balancing but Desirable but Conflicting Objectives in the Newsvendor Problem*, IIE Transactions, 35, 131-142.

Pinto R. (2016), *Stock Rationing under a Profit Satisficing Objective*, Omega, 65, 55-68.

Qin Y., Wang R., Vakharia A.J., Chen Y., M.H. Seref (2011), *The Newsvendor Problem: Review and Directions for Future Research*, European Journal of Operational Research, 213, 361-374.

Ray P., Jenamani M. (2016), *Mean-variance Analysis of Sourcing Decision under Disruption Risk*, European Journal of Operational Research, 250(2), 679-689.

Rubio-Herrero J., Baykal-Gursoy M., Jaśkiewicz A. (2015), *A Price Setting Newsvendor Problem under Mean-Variance Criteria*, European Journal of Operational Research, 247(2), 575-587.

Silver E.A., Pyke D.F., Peterson R. (1998), *Inventory Management and Production Planning and Scheduling*, Wiley, New York.

Wang C.X., Webster S. (2009), *The Loss Averse Newsvendor Problem*, Omega, 37, 93-105.

Wu M., Zhu S.X., Teunter R.H. (2014), *A Risk-averse Competitive Newsvendor Problem under the CVaR Criterion*, International Journal of Production Economics, 156, 13-23.

Wu M., Zhu S.X., Teunter R.H. (2013a), *The Risk-averse Newsvendor Problem with Random Capacity*, European Journal of Operational Research, 231(2), 328-336.

Wu M., Zhu S.X., Teunter R.H. (2013b), *Newsvendor Problem with Random Shortage Cost under a Risk Criterion*, International Journal of Production Economics, 145(2), 790-798.

Xinshenga X., Zhiqinga M., Ruia S., Mina J., Pingc J. (2015), *Optimal Decisions for the Loss-averse Newsvendor Problem under CVaR*, International Journal of Production Economics, 164, 146-159.

Yang S., Shi C.V., Zhao X. (2011), *Optimal Ordering and Pricing Decisions for a Target Oriented Newsvendor*, Omega, 39, 110-115.

Ye T., Sun H. (2016), *Price-setting Newsvendor with Strategic Consumers*, Omega, 63, 103-110.

M U L T I P L E   C R I T E R I A   D E C I S I O N   M A K I N G
Vol. 12 2017

Barbara Gładysz[*]
Dorota Kuchta[**]

MULTICRITERIA ANALYSIS OF THE SUCCESS OF RESEARCH PROJECTS

DOI: 10.22367/mcdm.2017.12.02

Abstract

The present paper considers the problem of the success of a research project evaluated by its outputs (which can be seen as the project's success measures) related to its inputs (constituting the project's success factors). Data Envelopment Analysis (DEA) models are used. The inputs and outputs are selected on the basis of a review of the literature. Two models are applied to a set of research projects implemented in Poland. Advantages and disadvantages of the approach are shown. In particular, the selection of inputs, outputs and their weights needs to be researched further. But the models used in the paper, in spite of their imperfection and lack of generality, considerably help to assess and compare the projects. Therefore, DEA is an important tool for the evaluation of R&D activities.

Keywords: research project, project success, Data Envelopment Analysis.

1 Introduction

Research projects consume a large amount of financial and human resources, provided by both state and business. At the same time, they often fail to bring the expected results. On the one hand, this is the nature of research projects: they explore previously unexplored domains and thus can lead to less significant results or even to no results, if their hypotheses turn out to be incorrect (Betta

[*] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wrocław, Poland, e-mail: barbara.gladysz@pwr.edu.pl.
[**] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wrocław, Poland, e-mail: dorota.kuchta@pwr.edu.pl.

et al., 2017; Kuchta et al., 2017; Gładysz and Kuchta, 2016). But it is still important to evaluate their outcomes or success as compared with the resources and effort invested, having in mind that the success of a research project should be sometimes measured in a different, less quantitative and more subjective way than that of, for example, a construction project (Chan et al., 2002). Without such evaluation the resources and efforts spent on R&D activities would be completely out of control.

A method which allows to compare inputs and outputs is the DEA method, which evaluates the efficiency of the so-called decision making units which may also be projects. The aim of the paper is to apply the DEA method to a set of research projects, using various sets of inputs and outputs, and to analyse the results critically, in order to assess the possibility of the applications of this method to the evaluation of research projects and to elaborate initial indicators as to the construction of a DEA model.

The set of research projects analysed here consists of research projects implemented in Poland in the last five years whose project managers filled out a questionnaire. We received 35 questionnaires completely filled out. The project managers were asked to select one of their research projects. Then they were asked about its features, the project team and resources invested and also about the project outcomes and their subjective evaluation.

We present the necessary theory of DEA and its state-of-the-art applications to research projects. Next, we describe the selected models and discuss the results. The paper concludes with remarks regarding the choice of inputs and outputs and their weights for DEA models applied to research project analysis.

2 Data Envelopment Analysis in research projects analysis

Data Envelopment Analysis (DEA) is a method which allows to determine, for a given set of objects (decision making units), those objects which are efficient. The efficiency assessment is conducted on the basis of the set of objects under consideration. The results are therefore not absolute, but relative efficiencies of the objects being examined. DEA is based on the productivity concept introduced by Debreu (1951) and Farrell (1957). The efficiency (which is in some cases a synonym for productivity) is measured as the relation of the effects (outputs) to the expenditures (inputs). The objects for which this relation is maximal (as compared to all the objects from the group under consideration) are assumed to be efficient. There are several DEA models (see Castelli et al., 2010). We adopt here the following one:

$$Max\ g_{i_0} = \frac{\sum_{r=1}^{R} \mu_{ri_0} y_{ri_0}}{\sum_{p=1}^{P} \gamma_{pi_0} x_{pi_0}} \tag{1}$$

with the constrains:

$$\frac{\sum_{r=1}^{R} \mu_{ri_0} y_{ri}}{\sum_{p=1}^{P} \gamma_{pi_0} x_{pi}} \leq 1$$

$\mu_{ri_0} \geq 0,\ r = 1, \ldots, R$
$\gamma_{pi_0} \geq 0\ p = 1, \ldots, P$
where:
$i = 1,\ldots, N$ – the index of an object,
$i_0 = 1, \ldots, N$ – the index of the object being evaluated at the moment,
x_{ri} for $r = 1, \ldots, R$, $i = 1,\ldots, N$ – the input values for the i-th object,
y_{pi} for $p = 1, .., P$, $i = 1,\ldots, N$ – the output values for the i-th object.

The maximal value of the objective function (1) is taken as the efficiency of the i_0-th object. As it can be concluded from this model, each object, when being evaluated, can choose for itself and for all the other objects the weights μ_{ri_0}, γ_{pi_0}, $r = 1, \ldots, R$ which show it in the best light.

The DEA method has been applied in the literature to project analysis. In that case the objects are projects. Here we focus on those literature items which deal with research projects. Research projects are projects which deal with Research and Development, which, in turn, "comprises creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society and the use of this stock of knowledge to devise new applications" (Frascati Manual, 2002). The following inputs for model (1) have been used in the literature for research projects (Eilat, Golany & Shtub, 2008; Revilla, Sarkis & Modrego, 2003; Yuan & Huang, 2002):

- project cost (budgeted and actual);
- full time equivalents of highly trained personnel (managers, engineers and scientists, PhD, master, bachelor degree holders) used for the realisation of the project;
- total revenues of the organisation;
- total R&D budget of the organisation, total number of its corporate employees;
- total number of the organisation's employees.

It has to be stressed that the last three inputs refer not to the project being evaluated, but to the whole organisation implementing the project.

As for the outputs, the following have been used for research projects (Eilat, Golany & Shtub, 2008; Revilla, Sarkis & Modrego, 2003; Yuan & Huang, 2002), many of them qualitative:

- discounted cash flow generated by the project;
- performance improvement achieved thanks to the project;
- customer satisfaction with the product of the project;
- congruence with the strategy of the organisation implementing the project;
- synergy with other projects realised by the organisation;
- project team satisfaction;
- the number of team members trained in project management thanks to the project realisation;
- probability of technological and commercial success of the project's product;
- the size of the technical gap filled by the project's product;
- the novelty of the technology used;
- the complexity of market activities needed to commercialize the project product;
- new employees gained thanks to the project;
- total income generated by the project;
- number of patents and copyrights gained thanks to the project;
- number of dissertations completed thanks to the project;
- number of reports issued thanks to the project;
- number of technology innovations designed thanks to the project;
- number of seminars organised thanks to the project;
- number of technology transfers resulting from the project.

Project outputs can serve as a measure of the project's success and project inputs as the factors of the project's success and failure. This is in line with the common belief (Betta et al., 2017) that the success or failure of research projects is not easily measurable and has to be based on several criteria. Apart from that, the notion of success or failure of a research project can be understood in different ways; success criteria may differ depending on the context. For example, if projects to be evaluated are implemented in an organisation whose main aim is to fulfil formal requirements of financing institutions, the measures of success would be, first of all, the relations between the planned and actual time and budget. However, if the organisation is a more open one, focusing on researching new, unexplored scientific questions and directions, the main measures of success would be the number of research questions answered (also negatively, with no tangible results) and of new research questions and partners identified.

DEA seems to be the right approach for assessing projects, including research projects. But we have to find out which inputs and outputs to choose in various situations. For our case study, we chose inputs and outputs using suggestions from the literature, but also trying to minimise the length and difficulty of the questionnaire which we distributed among research project managers in Poland. Our selection was meant to be a starting point for a discussion about the choice of inputs and outputs for DEA models applied to research projects. The model we used and the results are presented below.

3 Data Envelopment Analysis applied to a set of research projects

The DEA method was applied to 68 research projects implemented in Poland, in all disciplines. The data were gathered using on-line questionnaires. Although 68 questionnaires were filled out, only 35 were filled out completely; in 33 remaining cases several pieces of information, related mostly to the project's budget, were missing. The set of 35 completely filled out questionnaires constituted the basis for the study. Each questionnaire was filled out by a project manager of one research project, all of the respondents were university assistants or professors.

Both in the entire sample (68 projects) and in the subsample (complete questionnaires), all academic disciplines were represented, except for art (humanities 13%, social sciences 20%, pure sciences 17%, natural sciences 17%, technical sciences 12%, agricultural, forest and animal sciences 4%, medical sciences 17%).

Two DEA models were used. Out of all the potential inputs and outputs for research projects listed in the previous section, we selected those whose value could be inferred from the questionnaire, which had to be as short as possible. In both models the following inputs were used ($i = 1, 2, \ldots, 35$):
x_{1i} – the number of people in the project team,
x_{2i} – the actual duration of the project [months],
x_{3i} – the actual budget of the project [Polish currency PLN].

In the first model only quantitative outputs were selected:
y_{1i} – number of publications in journals indexed in the Master Journal List,
y_{2i} – number of publications in conference proceedings.

In the second model qualitative outputs were also taken into account; their values were assessed subjectively by the project manager:
y_{3i} – was the project successful?
y_{4i} – was the main project objective achieved?
y_{5i} – were the actual project results consistent with the initial assumptions?
y_{6i} – did the project implementation lead to unplanned results, important from the theoretical or practical point of view?

The values of the subjective outputs were evaluated on the Likert scale (1 – Strongly disagree, 2 – Disagree, 3 – Neither agree nor disagree, 4 – Agree, 5 – Strongly agree).

The descriptive statistics and histograms of inputs (x_{1i}, x_{2i}, x_{3i}) and outputs $(y_{1i}, \dots, y_{6i})$, i = 1, 2, …, 35, are presented in Table 1 and in Figures 1-3.

Table 1: Descriptive statistics for inputs and outputs

Descriptive Statistics	Input			Output					
				Quantitative		Qualitative			
	x_{1i} People	x_{2i} Duration	x_{3i} Budget	y_{1i} M.J.L.	y_{2i} Conf.	y_{3i} Success	y_{4i} Goal	y_{5i} Plan	y_{6i} Other results
Mean	5.9	36.3	395	4.9	4.9	4.4	4.6	4.3	4.3
St. Dev.	4.5	11.2	326	7.3	6.4	0.6	0.5	1.0	1.2
Min	1	12	39	0	0	3	3	1	1
Max	25	60	1500	40	35	5	5	5	5

The project teams consisted of 1-25 people, with mean cardinality about 6. The project's duration ranged from one to five years, with the average of about three years.

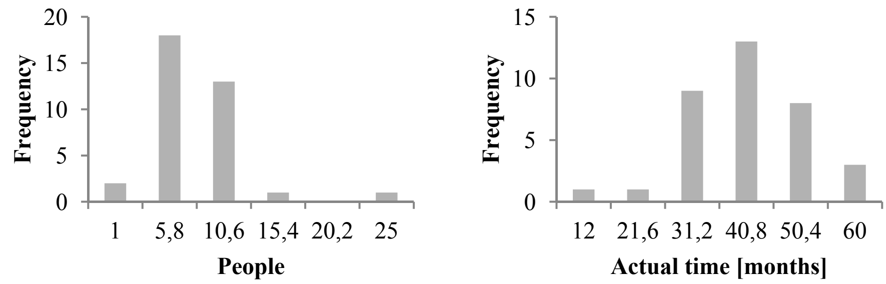


Figure 1. Histograms of inputs a): number of people in the project team and project duration (numbers represent interval endpoints)

The project budgets were between a few thousand and 1.5 million PLN, 400 thousand being the average.

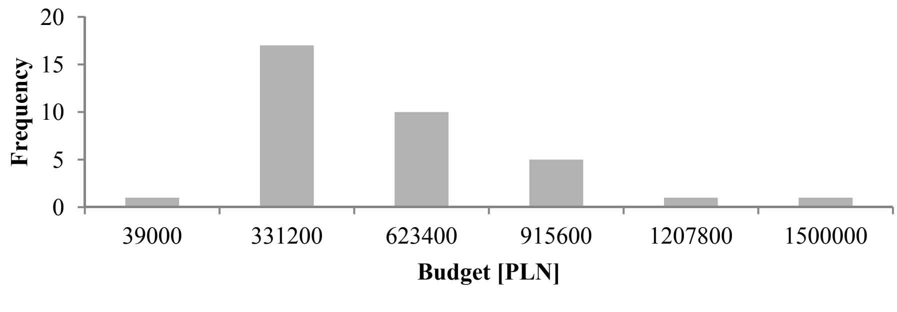


Figure 2. Histograms of inputs b): actual project budget (numbers represent interval endpoints)

The mean number of publications both in journals from the Master Journal List and in conference proceedings was about 5. However, the numbers of publications in journals from the Master Journal List were more diversified than those of publications in conference proceedings.

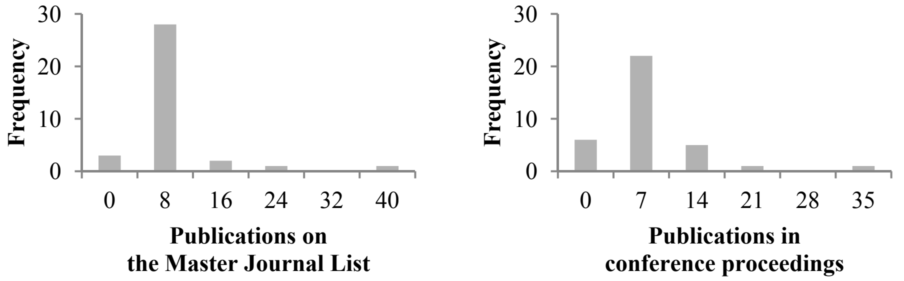


Figure 3. Histograms of quantitative outputs (numbers represent interval endpoints)

The mean qualitative assessment of four examined subjective measures of the research project's success is about 4.5. The assessments of the conformity of the project results with the initial plan and the attainment of unexpected but important results are more diversified than the other two assessments.

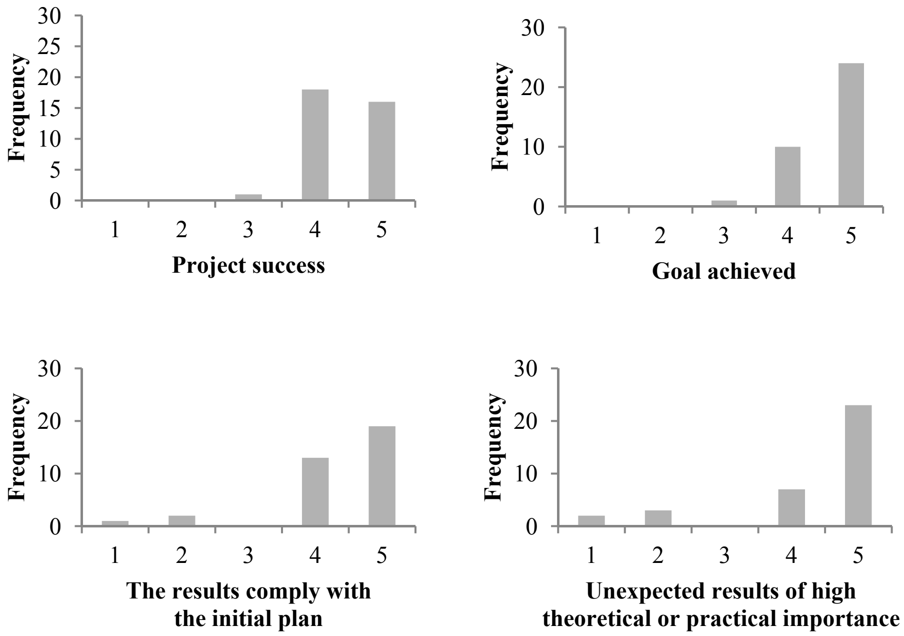


Figure 4. Histograms of qualitative outputs (numbers represent interval endpoints)

In the following section, the application of DEA to the set of projects described in this paper will be presented.

4 Results

Table 2 presents the results of the application of the DEA method using two versions of model (1): Model 1, with inputs (x_{1i}, x_{2i}, x_{3i}) and outputs (y_{1i}, y_{2i}), $i = 1, 2, \ldots, 35$, and Model 2, with the same inputs and with outputs $(y_{1i}, \ldots, y_{6i})$, $i = 1, 2, \ldots, 35$. In columns 11 and 12 the efficiencies calculated for the individual projects in both models are presented, in the last column the difference between the efficiency obtained in Model 2 and that calculated in Model 1 is shown.

Table 2: Inputs, outputs and project efficiencies

Project No. (*i*)	**Inputs**			**Outputs**						**Efficiency**		**Difference**
	x_{1i}	x_{2i}	x_{3i}	y_{1i}	y_{2i}	y_{3i}	y_{4i}	y_{5i}	y_{6i}	**Model 1**	**Model 2**	**M2-M1**
1	6	36	3000	0	2	5	5	5	4.3	0.10	0.65	0.55
2	2	24	100	5	0	5	5	5	1.2	1.00	1.00	0
3	3	30	108	0	7	3	4	4	1	0.92	1.00	0.08
4	3	27	50	5	3	5	5	4	5	1.00	1.00	0
5	5	13	39	0	2	4	5	4	5	0.73	1.00	0.27
6	7	36	205	5	0	4	3	2	5	0.47	0.61	0.14
7	3	24	99	3	2	4	5	5	1	0.51	1.00	0.49
8	10	36	400	9	7	5	5	5	5	0.56	0.69	0.13
9	2	26	68	1	4	4	5	4	4	0.04	1.00	0.96
10	4	48	750	7	4	5	5	5	5	0.84	0.75	-0.09
11	7	12	120	2	3	5	5	4	4	0.40	1.00	0.6
12	1	36	140	3	0	5	5	4	5	0.40	1.00	0.6
13	7	36	913	4	15	4	5	5	4	0.77	0.81	0.04
14	10	46	860	4	8	4	5	5	5	0.49	0.50	0.01
15	4	30	325	2	6	4	5	5	4	0.18	0.48	0.3
16	3	24	100	1	2	4	4	4	1	0.29	0.82	0.53
17	4	36	230	6	0	5	5	5	5	0.60	0.68	0.08
18	15	52	1500	40	0	4	4	4	2	1.00	1.00	0
19	3	48	336	1	0	4	4	4	5	0.10	0.50	0.4
20	7	36	335	3	3	5	5	5	2	0.21	0.51	0.3
21	1	26	60	1	3	4	4	2	4	0.71	1.00	0.29
22	5	36	516	3	11	4	4	5	5	0.50	0.60	0.1
23	25	57	440	2	1	5	5	4	5	0.10	0.25	0.15
24	4	24	100	2	1	5	5	5	4	0.33	0.88	0.55
25	10	36	460	3	5	5	5	5	5	0.17	0.41	0.24
26	5	48	600	3	12	4	4	4	5	0.55	0.53	-0.02
27	4	48	450	4	6	4	5	4	5	0.36	0.46	0.1
28	4	60	1000	2	10	4	4	1	5	0.57	0.15	-0.42
29	3	36	269	1	1	5	5	5	2	0.11	0.71	0.6
30	7	42	290	2	3	4	4	5	5	0.15	0.53	0.38
31	8	36	499	19	35	5	5	5	5	1.00	1.00	0
32	7	48	540	2	3	4	4	4	5	0.11	0.46	0.35
33	3	44	642	15	8	5	5	5	5	1.00	1.00	0
34	10	36	280	6	2	4	5	5	5	0.47	0.62	0.15
35	6	39	693	6	4	5	5	5	5	0.32	0.61	0.29

In the case of Model 1, in which only publications were counted as outputs, 13% of projects are effective to the degree 1. In the case of Model 2, where also qualitative (subjective) outputs were taken into account, 34% of projects were efficient to the degree 1. Here each project which was fully efficient in Model 1 was also fully efficient in Model 2. The distribution of project efficiencies are given in Table 3 and Figures 5 and 6.

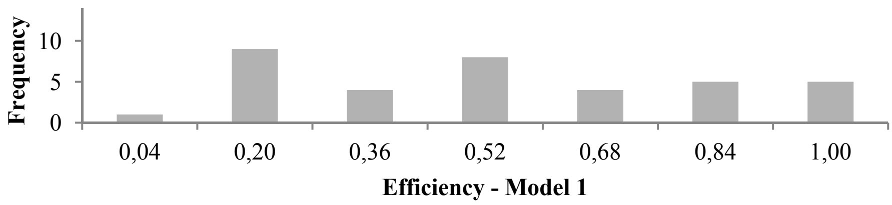


Figure 5. Histogram of project efficiencies for Model 1 (only quantitative outputs)

In the case of Model 1, although the best weights for each project can be chosen in (1), a large proportion of projects (about 54%) are clearly inefficient: their efficiency is under 0.5. The 13% of projects which turned out to be fully efficient are clearly better than most of the other projects. Only 20% of projects are characterised by an efficiency higher than 0.8.

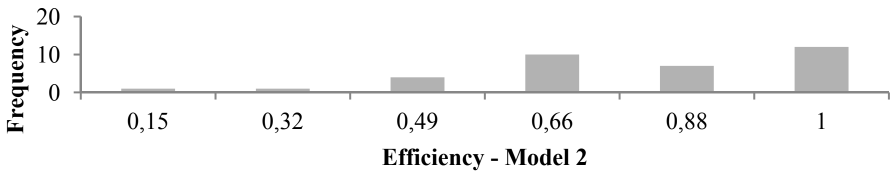


Figure 6. Histogram of project efficiencies for Model 2 (quantitative and qualitative outputs, numbers represent interval endpoints)

In the case of Model 2 the results are different. Over 40% of projects have proved to be efficient to the degree 0.8 or more and only less than 23% of projects are characterised by an efficiency smaller than 0.5. Of course, this is because adding more outputs often increases the project's efficiency with respect to Model 1. That is why in Model 2 the "picture" of the examined set of projects is much more favourable, which is visible also in Table 3.

Table 3: Efficiency of research projects

	Efficiency	
	Model 1	**Model 2**
Mean	0.48	0.72
St. Dev.	0.32	0.25
Min	0.04	0.15
Max	1	1

The results differ strongly between the two models. Thus, the question arises which model gives the “correct” information. As indicated in Introduction, in the literature on the applications of DEA to research projects, various inputs and outputs are considered, and therefore the issue is not clear. But the two models clearly point to a group of best projects (those which are fully efficient in Model 1) and to that of worst projects (those with an efficiency smaller than 0.4 in Model 2). These two groups can be analysed by the university management and treated as reference groups.

It is also interesting to look closer at projects whose efficiency varies considerably between the two models. For example, let us look closer at projects for which the absolute value of this difference is at least 0.6: projects 9, 11 and 12. All of them have a higher efficiency in Model 2.

- Project 9 had only one publication in a journal from the JCR basis, but all the qualitative assessments are equal to 4 or 5. The qualitative assessments were performed by the project manager, whose opinion may be subjective. But the project was not very expensive, had a small project team and was of medium duration. It is therefore possible that the importance assigned to the JCR basis unfairly underestimated the project and it is only thanks to the use of other outputs that the project was more fairly evaluated.
- Project 11 was assessed lower in Model 1 because it uses a comparatively large project team and budget. But again, the subjective assessment of the project manager considerably changed the assessment of the project.
- Project 12 was rather long and had no publications in conference proceedings, but the project manager assessed it highly. Hence the difference in the assessment of its efficiency in both models.

An interesting case is Project 28. It belongs to the small set of projects for which Model 2 found a lower efficiency than Model 1. This is because the project manager gave it a low mark in one of the subjective assessments, which was very rare in this set of projects. That is why its relative efficiency, when this low subjective assessment is taken into account (i.e. in Model 2), is rather low. This shows the dangers and problems related to subjective evaluations: the sensitivity of the results is very high in this aspect.

It is worth noting that four of the projects for which the assessment in Model 2 is much higher than in Model 1 belong to social sciences (projects 5, 9, 12 and 21). In these disciplines the importance attached to the publication quality is generally considered to be too high. Research in this field is often mainly of local interest and it is difficult to publish it in international journals. In this field, subjective assessments seem especially important. Moreover, most projects whose assessments in the two models do not differ very much, belong to pure, technical or medical sciences. Therefore it seems clear that project assessment should be conducted separately for various disciplines, as projects from different disciplines may be incomparable.

The final choice of outputs will depend on the definition of the success of a research project used by the organisations in question. If for the given organisation it is mainly the number of publications that counts (because of the national regulations in assessing research organisations), Model 1 or its modifications (for various groups of publications) will be used. If, on the other hand, the organisation has a less formal approach to the evaluation of research project success, trusts its project managers and wants to support also such projects to which not many publications can be linked directly, but which, in the opinion of the project managers, were efficient, it may want to adopt Model 2 or similar ones. This also depend on the field. Social sciences, for example, would require a more subjective model, while pure sciences, a more formal one.

It is worthwhile to emphasise the importance of subjective outputs. According to Chan and Chan (2004), project success (for projects generally) has to be evaluated using both objective and subjective criteria. As far as research projects are concerned, the research described in Betta et al. (2017) and based on about 70 interviews with research project managers in Poland and in France, clearly shows that research project managers criticise strongly such measures as the number and quality of publications and propose many other measures of research project success, most of them strongly subjective: new project ideas, new cooperation possibilities, providing an answer to an interesting research question etc. Most of the interviewed managers are convinced that any research project assessment ignoring subjective project outputs may distort project evaluation. It is thus important to include subjective outputs in DEA models for research projects.

In any case, using two models as different as Model 1 and Model 2 is also of some practical use, as small differences in the assessment point to those research projects whose assessment is more or less unequivocal. If we take the value 0.1 as the threshold for the absolute value of the difference from the last column of Table 2, we can see that the set of projects whose assessment was identical or almost identical in both models consists of projects 2, 4, 10, 13, 14, 17, 18, 22, 26, 27, 31, 33. For those projects, fairly conclusive statements can be made. We know to what extent they are efficient as compared with the other projects.

5 Conclusions

In our paper two DEA models have been applied to a set of research projects implemented in Poland. The models allowed to assess the projects using the relationships of project outputs (which can be regarded as measures of projects' success) and project inputs (which are measures of projects' relative efficiency). The method may help in the assessment of research projects.

However, the results have shown a high sensitivity of the results to the choice of inputs and outputs. We considered two models with the same set of inputs but different sets of outputs. The first set consisted of quantitative outputs only (the number of publications linked to the project, classified in various publication groups), the second set included the first one, but contained also qualitative, subjective outputs.

Our paper does not provide a final answer to the question which inputs and outputs have to be considered while assessing research projects. The problem is partially linked to the issue of defining the success of research projects, which is measured by means of its outputs. This is an open problem. One thing is clear: the understanding of a research project's success depends strongly on the stakeholder who decides about it (Davis, 2014); subjective outputs are at least as important as objective ones (Betta et al., 2017; Chan and Chan, 2004). Thus, the starting point in the selection of outputs should be an adequate definition of a project's success. Should it refer mainly to the number of publications resulting from the project and their quality or should it refer to the long-term potential future outputs of the project? In the latter approach, a project can be assessed as successful also when not many high-quality publications resulted from it, but it gave rise to new ideas or suggestions for further research. It is also clear that evaluation models should be built separately for various disciplines, since, for example, research in art or social sciences results in different outputs than that in pure or technical sciences.

As for inputs, they should reflect the most important expenditures (in a general sense of this notion) which the research organisation spends on research projects or its most scarce resources. They might be people (with various positions or degrees), time, money (quantitative inputs), but also expertise and experience in certain fields (qualitative inputs), laboratory space, work load etc. It may be that, for instance, the number of people in the project team should be modified to a weighted number of them, with the weight reflecting the irreplaceability and the experience of the individual team members. Also, the qualities of the project manager (also those expressed in qualitative terms) might constitute an input.

Another problem to be solved is related to the weights in model (1) which was the basis of our two models. In this model the weights can be freely changed to put the given project in the best light. This means that each project can "decide" which inputs and outputs are more important and the "decision" can be different for each project. But this approach may not be in line with the research institution's policy. If experienced researchers are scarce or if the publication on the Master Journal List counts for very much, it might be necessary to impose constraints on the corresponding weights. Such weights will not allow a project to overestimate its inputs and outputs which are only moderately important and underestimate the ones which are very important. This is line with the DEA approach, which allows for limits on weights (Podinovski, 2016).

In short, the DEA approach seems appropriate for the evaluation of research projects, but it needs more case studies in which project stakeholders and research institutions would participate. The aim of those case studies would be to determine inputs and outputs and the corresponding weights for model (1). There would possibly be various classes of situations (various strategies of the research institutions, various views of stakeholders) and for each different set of inputs, outputs and the corresponding weights would be generated. Once verified by numerous case studies, the DEA model for research projects might be a useful tool for assessing, comparing and then improving the efficiency of the research activities.

Acknowledgments

This work was partially supported by the National Science Centre in Poland "Czynniki sukcesu i porażki projektów badawczych. Studium przypadków Polski ("nowa" Unia) i Francji ("stara" Unia)", nb 2014/13/B/HS4/01660 and partially supported by the Polish Ministry of Science and Higher Education.

References

Betta J., Jastrzębska J., Frączkowski K., Gładysz B., Kuchta D., Prałat E., Ptaszyńska E., Rola P., Walecka-Jankowska K., Ropuszyńska-Surma E., Skomra A., Ryńca R., Klaus-Rosińska A., Mrzygłocka-Chojnacka J. (2017), *Success and Failure Factors of R&D Projects at Universities in Poland and France*, submitted to International Conference Business Risk in Changing Dynamics of Global Village April 26-27, 2017 at PWSZ in Nysa, Poland.

Castelli L., Pesenti R., Ukovich W. (2010), *A Classification of DEA Models when the Internal Structure of the Decision Making Units Is Considered*, Annals of Operations Research, 173, 207-235.

Chan A.P.C., Chan A.P.L. (2004), *Key Performance Indicators for Measuring Construction Success*, Benchmarking: An International Journal, 11(2), 203-221.

Chan A.P.C., Scott D. & Lam E.W.M. (2002), *Framework of Success Criteria for Design/Build Projects*, Journal of Management in Engineering, 18(3), 120-128.

Davis K. (2014), *Different Stakeholder Groups and Their Perceptions of Project Success*, International Journal of Project Management, 32(2), 189-201.

Debreu G. (1951), *The Coefficient of Recourse Utilisation*, Econometrica, 19(3), 273-292.

Eilat H., Golany B. & Shtub A. (2008), *Research Project Evaluation: An Integrated DEA and Balanced Scorecard Approach*, Omega, 36(5), 895-912.

Farrell M.J. (1957), *The Measurement of Productive Efficiency*, Journal of the Royal Statistical Society. Series A, 120(III), 253-281.

Frascati Manual (2002), *Proposed Standard Practice for Surveys on Research and Experimental Development*, 6th edition, www.oecd.org/sti/frascatimanual.

Gładysz B., Kuchta D. (2016), *Research Projects Success Factors*, Proceedings of EDULEARN16 Conference 4th-6th July 2016, Barcelona, Spain, 7591-7597.

Kuchta D., Gładysz B., Skowron D., Betta J. (2017), *Research Projects in Science Sector*, R&D Management, 47(1), 88-110.

Podinovski V.V. (2016), *Optimal Weights in DEA Models with Weight Restrictions*, European Journal of Operational Research, 254(3), 916-924.

Revilla E., Sarkis J. & Modrego A. (2003), *Evaluating Performance of Public-private Research Collaborations: A DEA Analysis*, Journal of the Operational Research Society, 54(2), 165-174.

Yuan B., Huang J.-N. (2002), *Applying Data Envelopment Analysis to Evaluate the Efficiency of R&D Projects − A Case Study of R&D in Energy Technology*, Technology Commercialization, 111-134.

M U L T I P L E   C R I T E R I A   D E C I S I O N   M A K I N G

Vol. 12 2017

Dorota Górecka[*]

BIPOLAR MIX – A METHOD FOR MIXED EVALUATIONS AND ITS APPLICATION TO THE RANKING OF EUROPEAN PROJECTS

DOI: 10.22367/mcdm.2017.12.03

Abstract

A great variety of multi-criteria decision aiding (MCDA) methods has already been developed but few papers have dealt with mixed data (qualitative and quantitative). MCDA techniques accepting different types of evaluations (such as deterministic, stochastic and/or fuzzy ones) are rather rare and not very well known, even though this issue is crucial from a practical point of view, since mixed evaluations occur very frequently in appraising and selecting projects and organizations, as well as in risk management modelling, among other fields.

This paper presents a new discrete MCDA tool developed for mixed performances of alternatives called BIPOLAR MIX. It is based on the classical BIPOLAR method proposed by Konarzewska-Gubała (1989), and on its modification, namely the BIPOLAR method with stochastic dominance (SD) rules, proposed by Górecka (2009). A numerical example at the end of the paper illustrates the problem of ordering projects applying for co-financing from the European Union (EU).

Keywords: decision analysis, MCDA, mixed data, BIPOLAR MIX, uncertainty modelling, European Union.

1 Introduction

The case of mixed data is not frequently discussed in the literature and MCDA methods accepting different types of evaluations (e.g. ordinal and cardinal as well as deterministic, stochastic and/or fuzzy) are rather rare and not very well known. Nev-

[*] Nicolaus Copernicus University in Toruń, Faculty of Economic Sciences and Management, Department of Econometrics and Statistics, Toruń, Poland, e-mail: dgorecka@umk.pl.

ertheless, this issue is vital from a practical point of view since mixed evaluations are frequent in real-life decision-making problems. They may occur:

- in appraising and selecting projects and organizations for various purposes,
- in the assessment of environmental impact,
- in establishing quality-of-living city rankings,
- in risk management modelling.

Examples of multi-criteria models that can be applied in such situations are:

- NAIADE (Munda, 1995; Munda et al., 1995),
- PAMSSEM (Martel et al., 1997; Guitouni et al., 1999),
- EVAMIX (Voogd, 1982; 1983),
- EVAMIX method with stochastic dominance rules (Górecka, 2010b; 2012),
- EVAMIX method for mixed evaluations (Chojnacka, Górecka, 2016).

Mixed evaluations have been also considered by Zaras (2004) and Ben Amor et al. (2007).

In some cases, though, these approaches are not well-suited for decision-making. Therefore, this paper presents a new discrete MCDA tool developed for mixed performances of alternatives, called BIPOLAR MIX. It is based on the classical BIPOLAR method proposed by Konarzewska-Gubała (1989) and on its modification, namely the BIPOLAR method with stochastic dominance (SD) rules, proposed by Górecka (2009).

The paper contains an introduction, three sections, and a conclusion. In Section 2, a general modelling framework is presented to clarify the context in which the proposed method can be applied. Section 3 presents the BIPOLAR MIX technique. Finally, Section 4 provides an example illustrating the problem of ordering projects which apply for co-financing from the European Union funds.

2 The context of problem modelling

Because of the very large amount of money channelled into the Regional Policy of the European Union (Cohesion Policy funding for 2014-2020 amounts to EUR 351.8 billion (www 1)), it is extremely important to allocate the financial means in the most effective way possible. That depends, among other things, on the appropriate choice of projects that are going to be co-financed. To help the decision-makers in this challenging task, MCDA methods, methods for making decisions in the presence of multiple, usually conflicting criteria, should be used, since the evaluation of the projects which apply for funding from the EU requires taking into account many different aspects: economic, financial, environmental, ecological, technical, technological, social and legal (Górecka, 2011; 2012).

The development of the BIPOLAR MIX method was driven by the distinctive features of the analysed decision-making problem, as well as the expectations and needs of the decision-makers engaged in the realisation of the EU Regional Policy, which are as follows (Górecka, 2011):

- the decision-making problem should be formulated as a problem of providing a complete order of the alternatives – it is essential for each applicant to be classified in the ranking and to know its own result (overall score), preferably a numerical one (points), since otherwise the results may be unconvincing for them;
- there is no room for the incomparability of the alternatives – the ranking should be complete as the argumentation that the project has not been selected for co-financing as incomparable with others will not be acknowledged by the applicants;
- the occurrence of ties in the ranking should be limited since this may create problems with dividing the funds;
- the problem is a group decision-making problem – experts involved in the appraisal of projects separately and independently evaluate a finite number of competing projects, and their diverse individual views must be incorporated into a joint final decision;
- there should be a possibility to employ both quantitative and qualitative criteria, and to use mixed data (deterministic and stochastic);
- decision-makers are able to reveal their preferences, but they do not have much time for the interaction and cooperation with the analyst;
- the possibility of the occurrence of complete compensation should be removed – in the case of some criteria it may be risky and in the case of others, projects should satisfy the so-called 'minimal quality';
- on the one hand, the decision aiding technique should not be too complicated so that decision-makers can explain to the applicants how it works and clarify the reasons for the rejection of their projects; on the other hand, the decision-making method should not be too simple to limit the possibilities of manipulating the results;
- it is desired that the decision aiding method allows us to determine whether the highly ranked projects are really good or just better than the weak ones.

The BIPOLAR MIX method responds to all these requirements (properties of the decision-making problem analysed and its participants)[1]. It is presented in the next section.

---

[1] Advantages and disadvantages of various MCDA techniques in the context of the European projects selection are presented in Górecka (2010b; 2011). Main strengths and weaknesses of selected MCDA approaches in the context of choosing a wedding venue are described in Górecka (2013).

3 The BIPOLAR MIX method

The BIPOLAR MIX technique is based on the BIPOLAR method (Konarzewska-Gubała, 1989; 1991) and on the modified BIPOLAR method with stochastic dominance rules (Górecka, 2009; 2010a; 2014a). As required, it allows us, among other things:

- to obtain a complete order of the alternatives;
- to use mixed data;
- to determine whether highly ranked alternatives are really good or just not bad – it allows for ranking and sorting alternatives as well as for determining their quality, taking into account what is good and undesirable from the decision-maker's point of view in the decision-making problem (reference system);
- to eliminate both the phenomenon of full compensation and the problem of the incomparability of the alternatives.

In this paper it is assumed that the performance of alternatives is given in a deterministic and stochastic way, and that the decision-maker(s) are risk-averse. Thus, if the evaluations are stochastic, we will use FSD/SSD[2] (see Quirk, Saposnik, 1962; Hadar, Russel, 1969) and AFSD/ASSD rules (see Leshno, Levy, 2002) for modelling preferences with respect to criteria measured on a cardinal scale, and OFSD/OSSD (see Spector et al., 1996) and OAFSD/OASSD rules (see Górecka, 2009; 2011; 2014c) for criteria measured on an ordinal scale.

We assume that when comparing alternatives a_i and a_l with respect to a single criterion, the following situations are distinguished: strict preference, weak preference and non-preference (see Roy, 1990; Górecka, 2009; 2014b; cf. Nowak, 2004; 2005):

- alternative a_i is strictly preferred to alternative a_l:

$$a_i \, P \, a_l \Leftrightarrow F_k^i \, SD \, F_k^l \quad \text{and} \quad \mu_k(a_i) - \mu_k(a_l) > p_k\,, \tag{1}$$

- alternative a_l is strictly preferred to alternative a_i:

$$a_l \, P \, a_i \Leftrightarrow F_k^l \, SD \, F_k^i \quad \text{and} \quad \mu_k(a_l) - \mu_k(a_i) > p_k\,, \tag{2}$$

- alternative a_i is weakly preferred to alternative a_l:

$$a_i \, Q \, a_l \Leftrightarrow F_k^i \, SD \, F_k^l \quad \text{and} \quad q_k < \mu_k(a_i) - \mu_k(a_l) \le p_k\,, \tag{3}$$

- alternative a_l is weakly preferred to alternative a_i:

$$a_l \, Q \, a_i \Leftrightarrow F_k^l \, SD \, F_k^i \quad \text{and} \quad q_k < \mu_k(a_l) - \mu_k(a_i) \le p_k\,, \tag{4}$$

[2] If a decision-maker has also a decreasing absolute risk aversion, then the TSD rule (see Whitmore, 1970) should be additionally applied. If a decision-maker is risk-seeking, then FSD/SISD/TISD1/TISD2 rules (see Goovaerts et al., 1984; Zaras, 1989) should be used.

- non-preference – otherwise,
  where:
- F_k^i, F_k^l – distribution of the evaluations of alternative a_i and alternative a_l, respectively, with respect to criterion f_k,
- SD – stochastic dominance relation: FSD/SSD/AFSD/ASSD or OFSD/OSSD/OAFSD/OASSD,
- $\mu_k(a_i)$, $\mu_k(a_l)$ – average performance (expected value of the distribution of the evaluations) of a_i and a_l, respectively, on criterion f_k,
- q_k – indifference threshold for criterion f_k,
- p_k – preference threshold for criterion f_k.

We assume that:

- $F = \{f_1, f_2,\ldots, f_n\}$ is a finite set of n examined criteria (it is assumed that all criteria are maximized),
- $A = \{a_1, a_2,\ldots, a_m\}$ is a finite set of m alternatives,
- $R = \{r_1, r_2,\ldots, r_r\}$ is a reference set consisted of two subsets:
  - $D = \{d_1, d_2,\ldots, d_d\}$ – 'good' reference alternatives,
  - $Z = \{z_1, z_2,\ldots, z_z\}$ – 'bad' reference alternatives,
  - $D \cup Z = R$, $D \cap Z = \emptyset$,
  - $\forall d_g \in D \quad \forall z_h \in Z \quad \forall k = 1,2,\ldots,n \quad \mu_k(d_g) \geq \mu_k(z_h)$,
  - $\forall d_g \in D \quad \forall z_h \in Z \quad \forall k = 1,2,\ldots,n \quad f_k(d_g) \geq f_k(z_h)$,

where:

- $\mu_k(d_g), \mu_k(z_h)$ – average performance (expected value) of reference alternatives d_g and z_h, respectively, on criterion f_k,
- $f_k(d_g), f_k(z_h)$ – performance of reference alternatives d_g and z_h, respectively, on criterion f_k.

The BIPOLAR MIX procedure consists of the following steps:

Step 1: Comparison of considered alternatives (a_i) with reference alternatives (r_j) to determine the decision-maker(s)' preference model

A. Calculation of aggregated preference index $c(a_i, r_j)$ for each pair (a_i, r_j), where $a_i \in A$, $r_j \in R$:

$$c(a_i, r_j) = \sum_{k=1}^{n} w_k \varphi_k(a_i, r_j) \tag{5}$$

where:

$$\varphi_k(a_i, r_j) = \begin{cases} 1, & \text{if} \quad F_k^i \; SD \, F_k^j \wedge \mu_k(a_i) - \mu_k(r_j) > p_k \\ -1, & \text{if} \quad F_k^j \; SD \, F_k^i \wedge \mu_k(r_j) - \mu_k(a_i) > p_k \\ \dfrac{\mu_k(a_i) - q_k - \mu_k(r_j)}{p_k - q_k}, & \text{if} \;\; F_k^i \; SD \, F_k^j \wedge q_k < \mu_k(a_i) - \mu_k(r_j) \le p_k \\ -\dfrac{\mu_k(r_j) - q_k - \mu_k(a_i)}{p_k - q_k}, & \text{if} \;\; F_k^j \; SD \, F_k^i \wedge q_k < \mu_k(r_j) - \mu_k(a_i) \le p_k \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

or:

$$\varphi_k(a_i, r_j) = \begin{cases} 1, & \text{if} \quad f_k(a_i) - f_k(r_j) > p_k \\ -1, & \text{if} \quad f_k(r_j) - f_k(a_i) > p_k \\ \dfrac{f_k(a_i) - q_k - f_k(r_j)}{p_k - q_k}, & \text{if} \quad q_k < f_k(a_i) - f_k(r_j) \le p_k \\ -\dfrac{f_k(r_j) - q_k - f_k(a_i)}{p_k - q_k}, & \text{if} \quad q_k < f_k(r_j) - f_k(a_i) \le p_k \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

depending on data, where:

- w_k – coefficient of importance for criterion f_k, $\sum_{k=1}^{n} w_k = 1$,
- F_k^i, F_k^j – distribution of the evaluations of alternative a_i and reference alternative r_j, respectively, with respect to criterion f_k,
- SD – stochastic dominance relation,
- $\mu_k(a_i), \mu_k(r_j)$ – average performance (expected value of the evaluations' distribution) of a_i and r_j, respectively, on criterion f_k,
- $f_k(a_i), f_k(r_j)$ – performance of alternative a_i and reference alternative r_j, respectively, on criterion f_k,
- q_k, p_k – indifference and preference thresholds for criterion f_k.

B. Calculation of credibility index $\omega(a_i, r_j)$ for each pair (a_i, r_j), where $a_i \in A$, $r_j \in R$:

$$\omega(a_i,r_j)=\begin{cases} c(a_i,r_j) & if \;\; c(a_i,r_j)>0 \;\; and \;\; \forall k \in I^- \;\; \mu_k(a_i)\geq v_k \\ c(a_i,r_j) & if \;\; c(a_i,r_j)<0 \;\; and \;\; \forall k \in I^+ \;\; \mu_k(r_j)\geq v_k \\ 0 & otherwise \end{cases} \tag{8}$$

or:

$$\omega(a_i,r_j)=\begin{cases} c(a_i,r_j) & if \;\; c(a_i,r_j)>0 \;\; and \;\; \forall k \in I^- \;\; f_k(a_i)\geq v_k \\ c(a_i,r_j) & if \;\; c(a_i,r_j)<0 \;\; and \;\; \forall k \in I^+ \;\; f_k(r_j)\geq v_k \\ 0 & otherwise \end{cases} \tag{9}$$

depending on data, where:

- $I^+(a_i,r_j)=\{k:\varphi_k(a_i,r_j)>0\}$,
- $I^-(a_i,r_j)=\{k:\varphi_k(a_i,r_j)<0\}$,
- v_k – veto threshold for criterion f_k.

Hypothesis 'a_i is preferred to r_j' is accepted when both the concordance and the non-discordance conditions are satisfied. The concordance condition is satisfied if aggregated preference index $c(a_i, r_j)$ is greater than 0, whereas the non-discordance condition is satisfied if $\forall k \in I^- \;\; \mu_k(a_i)\geq v_k$ or $f_k(a_i)\geq v_k$ (depending on data), where v_k is the lowest acceptable expected value of the distribution of the evaluations on criterion f_k or the lowest acceptable evaluation on criterion f_k (depending on data). Hypothesis 'r_j is preferred to a_i' is accepted if aggregated preference index $c(a_i,r_j)$ is smaller than 0 and $\forall k \in I^+ \;\; \mu_k(r_j)\geq v_k$ or $f_k(r_j)\geq v_k$ (depending on data). If the non-discordance condition is not satisfied and/or aggregated preference index $c(a_i, r_j)$ is equal to 0, both hypotheses are rejected.

Step 2: Determining the position of considered alternatives (a_i) in relation to the bipolar reference system (D,Z) and drawing final conclusions about them, i.e. preparing recommendations for the decision-maker(s)

A. Comparison of considered alternatives (a_i) with 'good' reference alternatives (d_g) from subset D – calculation of success index d_{iS} for each alternative a_i:

$$d_{iS}=\frac{1}{d}\sum_{g=1}^{d}\omega(a_i,d_g),\; d_{iS}\in[-1,1]. \tag{10}$$

Mono-sorting:

- category S1: alternatives a_i for which $d_{iS} > 0$ (type: *overgood*),
- category S2: alternatives a_i for which $d_{iS} = 0$,
- category S3: alternatives a_i for which $d_{iS} < 0$ (type: *undergood*).

Mono-ranking: according to the descending value of d_{iS}.

B. Comparison of considered alternatives (a_i) with 'bad' reference alternatives (z_h) from subset Z – calculation of anti-failure index d_{iN} for each alternative a_i:

$$d_{iN} = \frac{1}{z}\sum_{h=1}^{z}\omega(a_i, z_h)\,,\ d_{iN} \in [-1,1]. \tag{11}$$

Mono-sorting:

- category N1: alternatives a_i for which $d_{iN} > 0$ (type: *overbad*),
- category N2: alternatives a_i for which $d_{iN} = 0$,
- category N3: alternatives a_i for which $d_{iN} < 0$ (type: *underbad*).

Mono-ranking: according to the descending value of d_{iN}.

C. Cumulative assessment of considered alternatives (a_i) in terms of success achievement and failure avoidance – calculation of final score d_{iSN} for each alternative a_i:

$$d_{iSN} = \frac{d_{iS} + d_{iN}}{2},\ d_{iSN} \in [-1,1]. \tag{12}$$

Indices d_{iS} and d_{iN} induce two independent orders on the set of considered alternatives: a success-oriented one and an anti-failure-oriented one, respectively. Using both indices simultaneously we can rank and sort alternatives a_i bipolarly.

Bipolar-sorting:

- category B1: alternatives a_i for which $d_{iS} + d_{iN} > 0$ (type: *good*),
- category B2: alternatives a_i for which $d_{iS} + d_{iN} = 0$,
- category B3: alternatives a_i for which $d_{iS} + d_{iN} < 0$ (type: *bad*).

Bipolar-ranking: according to the descending value of d_{iSN}.

4 Illustrative example

This paper shows an application of the BIPOLAR MIX method to the mock-up process of appraising and ranking applications for financial aid from the European Regional Development Fund.

Sixteen infrastructure projects were considered. They concern the protection of surface waters, waste management and flood control, and include:

- construction and modernisation of wastewater and rainwater collection networks and wastewater treatment plants,
- implementation of a system of communal waste management, which includes the construction of sorting and composting plants and recultivation of landfills,
- modernisation of dikes.

These projects were evaluated using 11 criteria[3]: 1 deterministic (total cost) and 10 stochastic. Regarding the latter, five experts − specialists in environmental protection infrastructure − scored them[4] from 0 (the lowest evaluation) to 10 (the highest evaluation).

The model of preferences for the decision-making problem is presented in Table 1[5], while Table 2 provides the performance matrix for 16 projects from the case study and 4 reference projects (two 'good' and two 'bad'[6]). The results obtained using the BIPOLAR MIX method are shown in Table 3.

[3] The set of 11 criteria was constructed as follows: a list of the criteria (based, among other things, on the data available in the applications considered for project co-financing and on information from official documents related to the EU funds) was presented to five specialists on environmental protection infrastructure and European Union funds who could accept or reject each of them. They also had a possibility to add their own criteria to the preliminary list.

[4] To keep the classified data secret while allowing for objective evaluation, the descriptions of the projects were truncated and standardised.

[5] Weighting coefficients for evaluation criteria were established by the five experts on environmental protection infrastructure and EU funds with the help of the REMBRANDT system (Lootsma et al., 1990; Olson et al., 1995). The experts were also asked to determine values of indifference and preference thresholds for stochastic criteria. Two extreme opinions were disregarded and from the remaining three, the arithmetic mean was calculated. It was subsequently rounded to the nearest integer. Indifference and preference thresholds for the deterministic criterion (total cost) as well as veto thresholds for all criteria were set by the present author.

[6] The reference set was constructed by the present author. For stochastic criteria it was assumed that desirable performances of alternatives (experts' appraisal scores) are high (higher than or equal to 60% of points available, i.e. 6) and not too diversified, while undesirable performances are low and/or diversified. In the case of total cost (deterministic criterion) it was assumed that values less than or equal to PLN 5 million are desired, while values higher than or equal to PLN 20 million are undesired.

Table 1: Model of preferences

f_k	Criterion	Min/max	Type of data	w_k	q_k	p_k	v_k
f_1	Total cost [PLN million]	min	deterministic	0.12	1	3	30
f_2	Efficiency [0-10; 5 experts]	max	stochastic	0.19	1	3	3
f_3	Influence on environment [0-10; 5 experts]	max	stochastic	0.15	2	4	3
f_4	Influence on employment [0-10; 5 experts]	max	stochastic	0.05	3	4	2
f_5	Influence on inhabitants' health [0-10; 5 experts]	max	stochastic	0.14	3	5	2
f_6	Influence on investment attractiveness [0-10; 5 experts]	max	stochastic	0.07	2	4	2
f_7	Influence on tourist attractiveness [0-10; 5 experts]	max	stochastic	0.06	2	5	2
f_8	Validity of the technical solutions [0-10; 5 experts]	max	stochastic	0.08	1	3	2
f_9	Sustainability and institutional feasibility of the project [0-10; 5 experts]	max	stochastic	0.06	1	3	2
f_{10}	Complementarity with other projects [0-10; 5 experts]	max	stochastic	0.04	2	4	2
f_{11}	Comprehensiveness [0-10; 5 experts]	max	stochastic	0.04	2	4	2

Table 2: Performance matrix

f_k	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}
a_i	$f_k(a_i)$	$\mu_k(a_i)$									
a_1	8.42	5.6	7.2	4.4	4.8	4.6	4.6	7.8	7.4	7.0	4.8
a_2	31.55	7.2	9.2	7.8	5.8	7.8	9.0	8.4	8.4	8.6	9.0
a_3	9.24	7.0	8.4	3.8	5.0	5.8	6.2	7.6	6.4	3.4	4.0
a_4	9.25	7.6	8.8	7.2	5.6	7.6	8.4	8.4	7.4	3.8	4.6
a_5	5.93	5.8	8.4	7.4	5.8	7.2	4.6	8.2	8.6	7.4	5.6
a_6	20.00	6.8	6.6	8.4	6.0	7.6	6.8	6.6	9.0	7.0	5.4
a_7	26.01	4.6	7.6	6.2	6.0	6.6	6.6	8.2	8.0	7.4	5.8
a_8	5.85	7.0	8.4	5.4	7.0	5.8	6.4	8.6	8.2	6.2	5.6
a_9	5.6	7.0	7.4	3.6	5.6	5.0	5.0	7.6	7.6	8.6	5.8
a_{10}	7.00	6.0	8.0	4.0	5.6	6.8	7.2	8.4	8.0	8.4	6.2
a_{11}	6.22	5.4	7.6	3.2	5.0	5.6	6.2	7.2	7.4	8.4	7.4
a_{12}	33.95	6.2	7.8	6.2	5.8	5.4	4.4	7.6	7.0	8.4	7.4
a_{13}	7.00	7.2	9.0	5.6	7.8	6.4	6.2	8.8	7.0	6.2	6.8
a_{14}	13.87	6.8	8.0	7.2	6.8	4.6	4.0	7.8	7.0	4.0	7.8
a_{15}	10.53	6.6	7.4	7.8	6.8	6.4	7.0	7.8	7.6	3.8	6.8
a_{16}	9.02	9.0	7.2	1.0	6.0	6.2	6.8	9.2	8.6	8.6	5.4
r_j	$f_k(r_j)$	$\mu_k(r_j)$									
d_1	5.00	9.0	9.0	7.0	7.0	7.0	7.0	7.0	7.0	8.0	8.0
d_2	3.00	7.2	7.4	6.6	7.8	7.6	7.4	7.2	8.0	6.4	6.8
z_1	20.00	4.0	5.0	4.0	6.0	6.0	4.0	5.0	5.0	4.0	4.0
z_2	30.00	4.6	3.8	5.6	4.8	5.4	4.4	4.2	4.0	5.0	5.2

Table 3: Rankings of the projects

No.	Monorankings of the projects				Bipolar ranking of the projects	
	a_i	Success indices d_{iS}	a_i	Anti-failure indices d_{iN}	a_i	Final score d_{iSN}
1	a_{16}	-0.083	a_4	0.636	a_{13}	0.238
2	a_8	-0.084	a_{13}	0.576	a_8	0.236
3	a_{13}	-0.100	a_8	0.555	a_4	0.234
4	a_9	-0.120	a_{14}	0.525	a_9	0.199
5	a_2	-0.136	a_9	0.519	a_{10}	0.156
6	a_4	-0.169	a_3	0.510	a_5	0.144
7	a_5	-0.175	a_{10}	0.495	a_{15}	0.140
8	a_{10}	-0.183	a_{15}	0.493	a_{14}	0.138
9	a_6	-0.183	a_5	0.463	a_3	0.135
10	a_{15}	-0.213	a_{11}	0.396	a_6	0.100
11	a_{12}	-0.225	a_6	0.383	a_{16}	0.089
12	a_{11}	-0.230	a_1	0.317	a_{11}	0.083
13	a_3	-0.240	a_{16}	0.261	a_1	0.013
14	a_{14}	-0.250	a_7	0.256	a_7	-0.017
15	a_7	-0.289	a_2, a_{12}	0.000	a_2	-0.068
16	a_1	-0.292			a_{12}	-0.113

According to the analysis, all the projects in the case study belong to the category S3 (*undergood*) and none belongs to the category N3 (*underbad*). The final scores show that 13 projects were classified into category B1 (so-called 'good alternatives'), namely: a_{13}, a_8, a_4, a_9, a_{10}, a_5, a_{15}, a_{14}, a_3, a_6, a_{16}, a_{11} and a_1. Project a_{13} turned out to be the strongest and project a_8, second-strongest. The worst project for subsidising was a_{12}. This project, as well as a_2 and a_7, have been classified into category B3 (so-called 'bad alternatives') and should definitely not be recommended for co-financing. This is because a_2 and a_{12} are very expensive (they cost PLN 31.55 million and PLN 33.95 million, respectively), which was clearly caught by the BIPOLAR MIX method thanks to the veto procedure applied in this technique. Project a_7 scored low (it was almost always worse or not better than both 'good projects' and in many cases even not better than 'bad projects'). Moreover, it is also quite costly (PLN 26.01 million).

5 Conclusions

The BIPOLAR MIX method proposed in this paper is an efficient and fully operable technique that can enhance the European project evaluation procedure and improve the decision-making process since the existing procedure, based most frequently on the weighted sum, is not free of drawbacks (see Górecka, 2009; 2010a; 2010b; 2011). On the one hand, it is not too simple (to limit the temptation of manipulating the results), and, on the other hand, it is not too complicated (to enable decision-maker(s) to understand how it works). Furthermore, it allows us to use mixed information (ordinal and cardinal as well as deterministic and stochastic evaluations) and it eliminates both the possibility of full compensation and the problem of the incomparability of the alternatives. In addition, it allows us to rank and sort the alternatives and to determine their quality, using the reference system determined by the decision-maker(s). Finally, it allows us to obtain a numerical final score and it is not labour-intensive or time-consuming for the decision-makers.

The BIPOLAR MIX method can also be used to solve other decision-making problems, such as the evaluation and selection of public service organizations all over the world (cf. Chojnacka, Górecka, 2016). In the not-too-distant future we will apply it to charities operating in Poland and other countries, for instance, in Australia and Great Britain.

References

Ben Amor S., Jabeur K., Martel J.M. (2007), *Multiple Criteria Aggregation Procedure for Mixed Evaluations*, European Journal of Operational Research, 18(3), 1506-1515.

Chojnacka E., Górecka D. (2016), *Evaluating Public Benefit Organizations in Poland with the EVAMIX Method for Mixed Data*, Multiple Criteria Decision Making, 11, 36-50.

Goovaerts M.J., De Vylder F., Haezendonck J. (1984), *Insurance Premiums: Theory and Applications*, North-Holland, Amsterdam.

Górecka D. (2009), *Wielokryterialne wspomaganie wyboru projektów europejskich*, TNOiK „Dom Organizatora", Toruń.

Górecka D. (2010a), *Wykorzystanie metod wielokryterialnych w procesie oceny i wyboru wniosków o dofinansowanie realizacji projektu z funduszy Unii Europejskiej*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 108, 76-91.

Górecka D. (2010b), *Zastosowanie metod wielokryterialnych opartych na relacji przewyższania do oceny europejskich projektów inwestycyjnych* [in:] M. Nowak (red.), *Metody i zastosowania badań operacyjnych'10*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, 100-125.

Górecka D. (2011), *On the Choice of Method in Multi-criteria Decision Aiding Process Concerning European Projects* [in:] T. Trzaskalik, T. Wachowicz (eds.), *Multiple Criteria Decision Making '10-11*, Publisher of The University of Economics in Katowice, Katowice, 81-103.

Górecka D. (2012), *Sensitivity and Robustness Analysis of Solutions Obtained in the European Projects' Ranking Process* [in:] T. Trzaskalik, T. Wachowicz (eds.), *Multiple Criteria Decision Making'12*, Publisher of the University of Economics in Katowice, Katowice, 86-111.

Górecka D. (2012), *Applying Multi-Criteria Decision Aiding Techniques in the Process of Project Management within the Wedding Planning Business*, Operations Research and Decisions, 22(4), 41-67.

Górecka D. (2014a), *Metoda BIPOLAR z dominacjami stochastycznymi* [in:] T. Trzaskalik (red.), *Wielokryterialne wspomaganie decyzji. Metody i zastosowania*, PWE, Warszawa, 149-152.

Górecka D. (2014b), *Metoda PROMETHEE II z progami weta i dominacjami stochastycznymi* [in:] T. Trzaskalik (red.), *Wielokryterialne wspomaganie decyzji. Metody i zastosowania*, PWE, Warszawa, 122-124.

Górecka D. (2014c), *Reguły wyboru oparte na relacji prawie dominacji stochastycznej dla kryteriów ocenianych w skali porządkowej* [in:] T. Trzaskalik (red.), *Wielokryterialne wspomaganie decyzji. Metody i zastosowania*, PWE, Warszawa, 31-32.

Guitouni A., Martel J.-M., Bélanger M., Hunter C. (1999), *Managing a Decision Making Situation in the Context of the Canadian Airspace Protection*, Working paper 1999-021, F.S.A., Université Laval, Canada.

Hadar J., Russel W. (1969), *Rules for Ordering Uncertain Prospects*, American Economic Review, 59, 25-34.

Konarzewska-Gubała E. (1989), *Bipolar: Multiple Criteria Decision Aid Using the Bipolar Reference System*, Cahiers et Documents du LAMSADE, Université Paris IX, Paris.

Konarzewska-Gubała E. (1991), *Wspomaganie decyzji wielokryterialnych: system BIPOLAR*, Wydawnictwo Uczelniane Akademii Ekonomicznej we Wrocławiu, Wrocław.

Leshno M., Levy H. (2002), *Preferred by "All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance*, Management Science, 48, 1074-1085.

Lootsma F.A., Mensch T.C.A., Vos F.A. (1990), *Multi-criteria Analysis and Budget Reallocation in Long-term Research Planning*, "European Journal of Operational Research", 47, 293-305.

Martel J.-M., Kiss L.R., Rousseau A. (1997), *PAMSSEM: Procédure d'agrégation multicritère de type surclassement de synthèse pour évaluations mixtes*, Manuscript, F.S.A., Université Laval.

Munda G. (1995), *Multicriteria Evaluation in a Fuzzy Environment*, Physica-Verlag, Heidelberg.

Munda G., Nijkamp P., Rietveld P. (1995), *Qualitative Multicriteria Methods for Fuzzy Evaluation Problems: An Illustration of Economic-ecological Evaluation*, European Journal of Operational Research, 82(1), 79-97.

Nowak M. (2004), *Preference and Veto Thresholds in Multicriteria Analysis Based on Stochastic Dominance*, European Journal of Operational Research, 158(2), 339-350.

Nowak M. (2005), *Investment Project Evaluation by Simulation and Multiple Criteria Decision Aiding Procedure*, Journal of Civil Engineering and Management, 11(3), 193-202.

Olson D.L., Fliedner G., Currie K. (1995), *Comparison of the REMBRANDT System with Analytic Hierarchy Process*, "European Journal of Operational Research", 82, 522-531.

Quirk J.P., Saposnik R. (1962), *Admissibility and Measurable Utility Functions*, Review of Economic Studies, 29, 140-146.

Roy B. (1990), *Wielokryterialne wspomaganie decyzji*, Wydawnictwa Naukowo-Techniczne, Warszawa.

Spector Y., Leshno M., Ben Horin M. (1996), *Stochastic Dominance in an Ordinal World*, European Journal of Operational Research, 93, 620-627.

Voogd H. (1982), *Multicriteria Evaluation with Mixed Qualitative and Quantitative Data*, Environment and Planning B, 9, 221-236.

Voogd H. (1983), *Multicriteria Evaluation for Urban and Regional Planning*, Pion, London.

Whitmore G.A. (1970), *Third-Degree Stochastic Dominance*, American Economic Review, 60, 457-459.

Zaras K. (1989), *Les dominances stochastiques pour deux classes de fonction d'utilité: concaves et convexes*, RAIRO/RO, 23, 57-65.

Zaras K. (2004), *Rough Approximation of a Preference Relation by a Multi-attribute Dominance for Deterministic, Stochastic and Fuzzy Decision Problems*, European Journal of Operational Research, 159(1), 196-206.

(www 1) European Commission: Regional Policy – Inforegio, http://ec.europa.eu/regional_policy/en/ (accessed: 1.06.2017).

MULTIPLE CRITERIA DECISION MAKING

Vol. 12 2017

Przemysław Juszczuk[*]
Ignacy Kaliszewski[**]
Janusz Miroforidis[***]

TRADE-OFF GUIDED SEARCH FOR APPROXIMATE PARETO OPTIMAL PORTFOLIOS

DOI: 10.22367/mcdm.2017.12.04

Abstract

In this paper, we attempt to represent the Pareto Front in the Markowitz mean-variance model by two-sided discrete approximations. We discuss the possibility of using such approximations for portfolio selection. The potential of the approach is illustrated by the results of preliminary numerical experiments.

Keywords: portfolio optimization, Pareto front approximation, Pareto front navigation.

1 Introduction

The standard approach to solving the Markowitz mean-variance model Markowitz (1952), Markowitz (1991), Gondzio et al. (2007), Elton et al. (2014) is to solve a number of optimization problems with a quadratic objective function, representing portfolio variance, and one linear function, representing portfolio mean return, constrained to be equal to a specific value. By this, a number of efficient (in the sense of Pareto) portfolios and the corresponding pairs of

[*] University of Economics in Katowice, Faculty of Informatics and Communication, Department of Knowledge Engineering, e-mail: przemyslaw.juszczuk@ue.katowice.pl.

[**] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, e-mail: ignacy.kaliszewski@ibspan.waw.pl.

[***] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, e-mail: jmiroforidis@gmail.com.

mean return and variance values (elements of the Pareto Front) can be derived. We refer to the mean-variance model, because the vast majority of research this subject is focused on this particular problem, rather than on the mean-standard deviation model. The mean-variance model has been a subject of extensive investigations which focused mostly on additional constraints included in the model such as skewness Briec et al. (2013), liquidity Lo et al. (2003) or portfolio size Chiam et al. (2008). Examples of large-scale portfolio optimization problems were discussed in Steuer et al. (2011) and details related to the large-scale optimization with multiple objectives are discussed in Steuer et al. (2006). To solve large-scale portfolio optimization problems, customized evolutionary algorithms were proposed, e.g., in Chen et al. (2011) and Deb et al. (2011).

This standard approach, however completely overshadows the inherent multiple criteria (here: bi-criteria) nature of the problem, where an investor looking for a portfolio he/she would prefer the most, trades risk (captured as variance) for gains (captured as mean returns) and vice versa.

Here we aim to assist investors which invest in portfolios on any scale, with a specific focus on investing in a large number of assets. We start with the observation that investors, with a possible exception of complete novices in the business, have some, maybe vague, idea about their individual risk profiles, for instance: being risk-prone, risk-neutral or risk-averse. This roughly narrows their potential investments to specific segments of the Pareto Front (PF). Hence, this makes the derivation of the whole PF superflous. To enable investors to locate such investor risk-specific segments, we propose a framework to search for the preferable combination of mean return and variance.

The framework consists in capturing the investors' risk profiles as they are willing to make on the unachievable ideal portfolio of zero risk and the maximal mean return to get an efficient portfolio. In the mean-variance model, the maximal mean return is yielded by investing the whole capital into the asset of the highest return.

To avoid having to solve optimization problems at the early stage of the decision making process, instead of the element of the PF of the required risk and mean return, investors can be provided with lower and upper bounds on each of these values. Bounds, if sufficiently tight, are used to decide if portfolios with risk and mean return within those bounds are satisfactory. If so, an element of the PF (and the corresponding portfolio) is derived by solving one quadratic optimization problem.

The outline of this paper is as follows. In Section 2, we recall the Markowitz model and provide necessary preliminaries. In Section 3, we show how the investor can be supported in his/her portfolio selection decisions by approximate valuations of elements of the PF in regions of that set of his/her temporary interest. In Section 4 and Section 5, we show how to populate two specific sets necessary to provide approximate valuations. Section 6 provides some illustrative numerical examples, whereas Section 7 concludes the paper.

2 Preliminaries

The problem of portfolio investments is formulated as follows. Given a number of risky assets, find a portfolio with the most preferred risk and return characteristic.

The underlying model for that problem is the Markowitz mean-variance (MV) model:

$$\begin{aligned} &\min\ f_1(x) = x^T Q x \quad \text{(minimize variance)} \\ &\max\ f_2(x) = e^T x \qquad \text{(maximize mean)} \end{aligned} \tag{1}$$

$$\text{subject to } x \in X_0 = \left\{ x \;\middle|\; \begin{array}{l} u^T x = 1, \ \text{(all capital to be consumed)}, \\ x \geq 0, \end{array} \right\},$$

where x is the vector of fractions of the capital, spent on buying individual assets, Q is the covariance matrix, e is the vector of means, u is the all-ones vector, x, u, $e \in R^n$, Q is an $n \times n$ matrix, and n is the number of assets.

Below we use the standard definition of solution (here: portfolio) efficiency in the sense of Pareto, and we refer to the set of valuations of efficient portfolios as the Pareto Front.

3 Selecting an investment portfolio

The method of preference-driven navigation over the Pareto front (PF), as proposed in Kaliszewski (2006), relies on the notion of *the vector of concessions*, serving as *the preference carrier*.

In what follows, we make use of an element y^* of R^2, $y_1^* = 0$, $y_2^* = \max_{x \in X_0} e^T x + \varepsilon$, $\varepsilon > 0$. Element y^* clearly does not represent any portfolio composed of risky assets.

Since y^* is unattainable, to get a feasible portfolio represented on the PF, the investor has to compromise on y^*, and he or she can do this by selecting a vector of concessions (see Kaliszewski et al., 2016) τ, $\tau_1 < 0, \tau_2 > 0$.

Vector τ defines proportions in which the investor agrees to sacrifice unattainable values of risk and mean return represented by y^* in a quest for an element of the PF.

Vector τ can be defined by the investor explicitly, in the *atomistic* (or *parametric*) *manner*, by indicating its components τ_1, τ_2 (in absolute numbers, e.g. $\tau = (-2, 7)$ or relative quantities, e.g. $\tau = (-\frac{2}{2+4}, \frac{7}{2+4})$), or implicitly, in the *holistic manner*, by indicating a *base point* y, i.e. a variant from the set $\{y \mid y_1 \in y_1^* + R_+, y_2 \in y_2^* - R_+\}$, R_+ – the set of real positive numbers, which defines τ as $\tau_1 = y_1 - y_1^*$, $\tau_2 = y_2^* - y_2$. The latter manner is a special version of the reference point paradigm (cf. e.g. Kaliszewski, 2006; Ehrgott, 2005; Miettinen, 1999; Wierzbicki, 1999).

If the investor expresses his or her preferences in the form of vectors τ, then for any such a vector, he or she can be provided with bounds:

$$L_l(S_L, \tau) \leq f_l(x^\tau) \leq U_l(VS_U, \tau), \; l = 1, 2, \tag{2}$$

where x^τ would be a solution to:

$$\min_{x \in X_0} \max\{\frac{1}{\tau_1}(f_1(x) - y_1^*), \frac{1}{\tau_2}(y_2^* - f_2(x))\} \tag{3}$$

if this problem were solved to optimality, S_L is a set of feasible portfolios, VS_U is a set of elements of R^2 located "above" the PF (for definitions of S_L and VS_U and formulas for $L_l(S_L, \tau)$ and $U_l(VS_U, \tau)$ see Miroforidis (2010), Kaliszewski et al. (2009), Kaliszewski et al. (2010), Kaliszewski et al. (2012a)). Figure 1 illustrates the derivation of bounds for given τ and unknown x^τ.

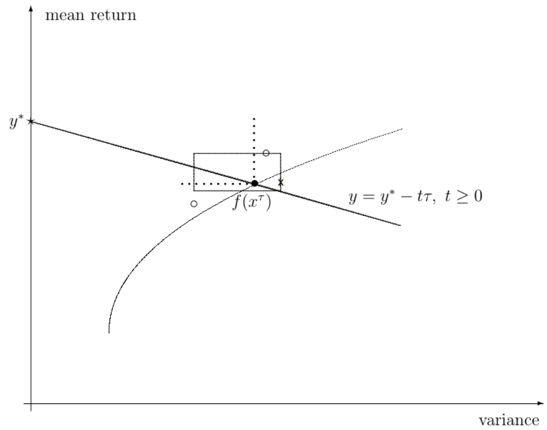


Figure 1. Derivation of bounds for given τ and unknown x^τ (bullet) with VS_U consisting of two elements (circles) and with S_L consisting of one portfolio (marked with an X). The North-West corner of the rectangle determined with these three elements is a lower bound for $f_1(x^\tau)$ and an upper bound for $f_2(x^\tau)$, whereas the South-East corner is an upper bound for $f_1(x^\tau)$ and a lower bound for $f_2(x^\tau)$

4 Derivation of S_L

To ensure effectiveness of lower bound calculations, sets S_L (termed *lower shells* in Miroforidis (2010), Kaliszewski et al. (2009), Kaliszewski et al. (2010), Kaliszewski et al. (2012), Kaliszewski et al. (2012a)) should be composed of elements which are not dominated by any other element in this set.

To populate S_L, we propose the following procedure:

Procedure 1: Derivation of the lower approximation S_L

```
begin
1   Initialize A with assets
2   for a number of pairs of portfolios in A do
3       For each pair derive a number of their convex combinations
4   S_L := A;
5   Delete the dominated portfolios from S_L; go to Step 2.
```

If the lower bounds obtained with this population procedure are not satisfactory, i.e. the differences between the lower and upper bounds are not as small as required, the procedure can be extended to combinations of more than two portfolios.

5 Derivation of VS_U

Similarly to set S_L, to ensure effectiveness of upper bound calculations, sets VS_U (termed *virtual upper shells* in Kaliszewski et al. (2012)) should be composed of elements which do not dominate any other element in this set.

Let us consider the following modification of model (1):

$$\begin{aligned} &\min\ f_1(x) = x^T Q x \\ &\max\ f_2'(x) = \tfrac{1}{\alpha}\, e^T x \end{aligned} \tag{4}$$

$$\text{subject to } x \in X_0\,.$$

Figure 2 represents PFs of this model for different α. It is clearly seen that for $\alpha = 1$ the PF to model (4) coincides with the PF of model (1) and for $\alpha < 1$ the PF satisfies the requirements for set VS_U for model (1).

Assume for the moment that some algorithm applied to model (1) has produced the PF and hence to model (4) with $\alpha = 1$. Then a simple rescaling of that PF in the model (4) with $\alpha < 1$ produces the PF to model (1) and hence VS_U (Figure 2).

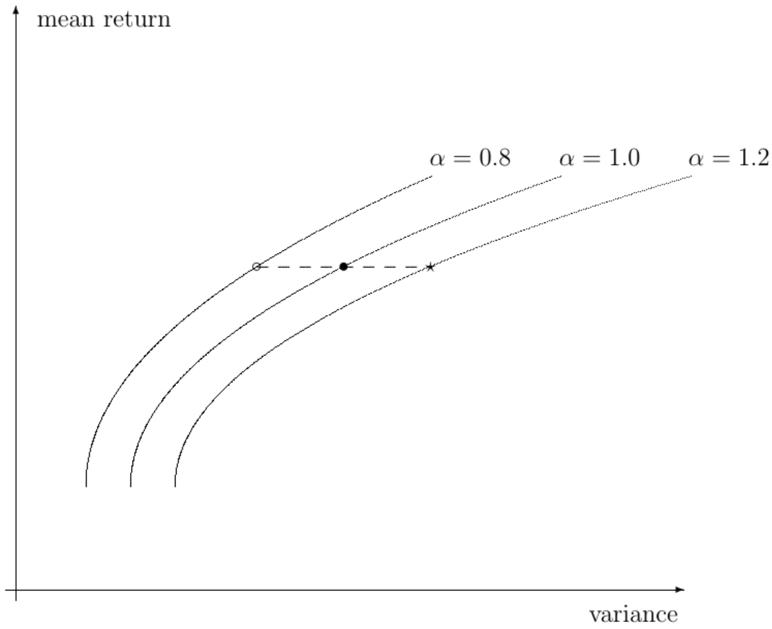


Figure 2. Pareto Fronts for model (4) with different α

This observation is of no practical value, because the PF for model (1) is the very object sought, not a given one. However, below we make use of the idea of objects shifting in the mean-variance space along the variance axis to populate VS_U.

Suppose that an algorithm is able produce S_L such that $\max_{x \in S_L}(f_1(x) - f_1(x')) \leq \beta$, where $f(x')$ are elements of the PF, such that for each $x \in S_L$, $f_1(x) = f_1(x')$.

Then, the set $\{y \,|\, y_1 = f_1(x) - \beta, y_2 = f_2(x),\ x \in S_L\}$ is clearly a valid VS_U.

At the moment we are not in the position to propose any exact method to derive β except in-sample testing. In the next section, we present results of a few such tests.

6 An illustrative example

We illustrate the idea on an example from the Beasley OR Library (1991), namely the problem *port4.txt* with 98 assets. For that problem the library provides the Pareto Front with 2000 elements uniformly covering the range of accessible returns (Figure 3).

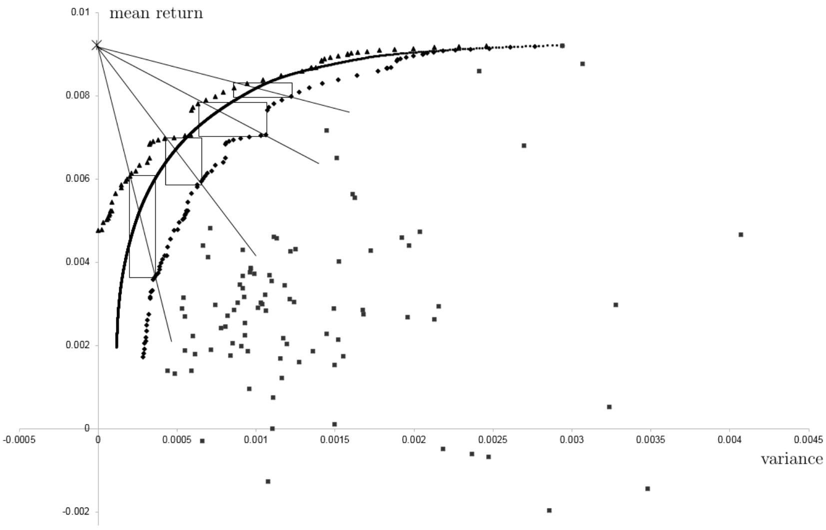


Figure 3. The first iteration – VS_U (▲), $f(S_L)$ (♦) and images of assets (■) in the example problem. For each τ and the corresponding compromise half line, the North-West corners of the rectangles represent the lower bound for variance and the upper bound for mean, and the South-East corners represent the upper bound for variance and the lower bound for mean

For this problem we have derived lower shell S_L by forming portfolios from pairs of assets (Figure 3). Next, for each portfolio from S_L we have calculated the difference in variance between this portfolio and an element of the Pareto Front with the same return. If there has been no element of the Pareto Front with the return equal to that of the portfolio, we have taken the element of the closest return. With 2000 elements of the Pareto Front, we miss the correct value of return by at most $0.1815 \cdot 10^{-5}$. Next, we shifted S_L along the horizontal axis by the value of the maximal difference and we have obtained a valid VS_U. With S_L and VS_U in place, we have calculated bounds on $f_l(x^\tau),\ l = 1, 2$, as in Table 1.

Table 1: Lower and upper bounds on components of $f(x^\tau)$ for selected τ, first iteration

$\boldsymbol{\tau}$	$\boldsymbol{L_1(S_L, \tau)}$	$\boldsymbol{L_2(S_L, \tau)}$	$\boldsymbol{U_1(VS_U, \tau)}$	$\boldsymbol{U_2(VS_U, \tau)}$
(1,1)	0.00123	0.00798	0.00086	0.00828
(0.5,1)	0.00107	0.00706	0.00063	0.00788
(0.2,1)	0.00065	0.00593	0.00042	0.00699
(0.0667,1)	0.00037	0.00367	0.00020	0.00613

In the second iteration, we have derived S_L by forming portfolios from pairs of portfolios in S_L of the first iteration and calculated bounds again. The bounds are given in Table 2.

Table 2: Lower and upper bounds on components of $f(x^\tau)$ for selected τ, second iteration

$\boldsymbol{\tau}$	$\boldsymbol{L_1(S_L,\tau)}$	$\boldsymbol{L_2(S_L,\tau)}$	$\boldsymbol{U_1(VS_U,\tau)}$	$\boldsymbol{U_2(VS_U,\tau)}$
(1,1)	0.00090	0.00811	0.00100	0.00821
(0.5,1)	0.00074	0.00752	0.00080	0.00773
(0.2,1)	0.00047	0.00641	0.00054	0.00669
(0.0667,1)	0.00025	0.00452	0.00031	0.00537

Table 3 presents maximal errors which occur when taking $L(S_L,\tau)$ to represent $f(x^\tau)$, defined as:

$$err_l = 100 \cdot \frac{U_l(VS_U,\tau) - L_l(S_L,\tau)}{L_l(S_L,\tau)}, \; l = 1,2$$

or:

$$\overline{err}_l = 100 \cdot \frac{f(x^\tau) - L_l(S_L,\tau)}{L_l(S_L,\tau)}, \; l = 1,2,$$

with $f(x^\tau)$ approximated by solving problem (3) over the discrete approximation of the Pareto Front, as provided in Beasley OR Library for that problem. The errors have been calculated for the first and second iteration.

Table 3: Maximal relative errors when taking $L(S_L,\tau)$ to represent $f(x^\tau)$

	First iteration				**Second iteration**			
τ	err_1	err_2	$\overline{err}_1$	$\overline{err}_2$	err_1	err_2	$\overline{err}_1$	$\overline{err}_2$
	%				%			
(1,1)	43.01	3.83	18.33	2.69	10.09	1.16	8.00	0.27
(0.5,1)	69.81	11.64	23.02	8.35	11.63	2.13	5.83	1.10
(0.2,1)	54.51	17.78	17.97	13.07	11.68	4.65	0.96	4.16
(0.0667,1)	83.27	66.82	32.33	41.92	22.69	18.36	5.00	14.89

The numbers in Table 3 illustrate the phenomenon of fast improvement of approximations of the Pareto Front by lower shells. As in the first iteration, the relative errors are absolutely unacceptable, in the second iteration they drop to the level which, if still unacceptable, makes sense to proceed to the third and possibly successive iterations. And it should be stressed that this is only by taking pairwise combinations of portfolios[1].

[1] A similar behavior was observed for the other portfolio selection problems from the Beasley OR Library; due to limited space we confined ourselves to presenting numerical results for one problem only.

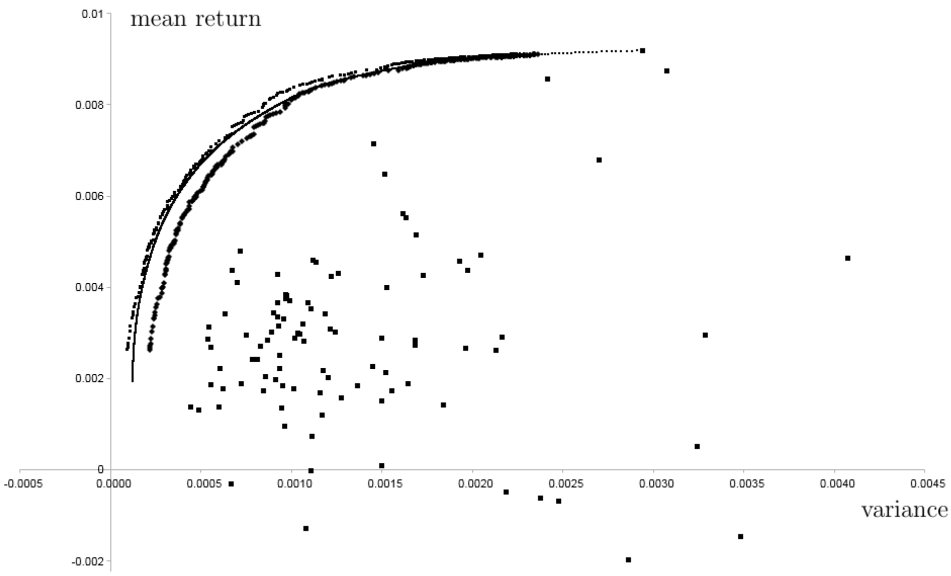


Figure 4. The second iteration – VS_U (▲), $f(S_L)$ (♦) and images of assets (■) in the example problem

Thus far we have attempted to approximate the entire Pareto Front, which is of limited use in practical applications. As we now know where to improve approximations locally (i.e. in the regions pointed to by the DM's preferences represented by the vectors of concessions and the corresponding compromise half line), we can limit computations solely to those regions.

7 Conclusions

In this paper, we have attempted to apply the general bounding methodology to the classical Markowitz portfolio selection mean-variance model. The methodology allows investors to express their preferences in a natural manner with the help of *vectors of concessions* and thus limit the search for Pareto optimal (efficient) portfolios directly to the regions of investors' interests. The existence of two-sided bounds on Pareto suboptimal portfolios allows to control the extent of Pareto suboptimality of feasible portfolios when they are considered for the most preferred portfolio.

We perceive inexact approaches to the portfolio selection problems to be a valid alternative to exact methods when the number of assets available for a portfolio exceeds a thousand. We raise three issues to support our view. First, inexact methods can provide feasible portfolios relatively quickly.

The only problem is their accuracy, but we have solved that problem by providing lower and upper bounds on portfolio variance and mean. Second, inexact methods are generally much simpler to code than exact methods, so they can be often coded in-house. This eliminates the need to acquire (often on the basis of a costly license) an exact solver. Third, in the case of problems admitting more constraints (e.g., cardinality constraints), exact methods become inefficient as the size and complexity of portfolio selection problems grows.

We have shown that in the case of the mean-variance portfolio selection problem, portfolios with a limited number of assets can provide reasonable approximations of the Pareto optimal portfolios. This observation needs to be further verified on various large-scale test problems. Applications of this observation to more complex portfolio selection problems will be the subject of the authors' further investigations.

References

Beasley J.E. OR-Library (1991), http://people.brunel.ac.uk/~mastjjb/jeb/info.html (10 July 2017).

Briec W., Kerstens K., van de Woestynec I. (2013), *Portfolio Selection with Skewness: A Comparison of Methods and a Generalized One Fund Result*, European Journal of Operational Research, 230(2), 412-421.

Chen Y., Mabu S., Hirasawa K. (2011), *General Relation Algorithm with Guided Mutation for the Large-scale Portfolio Optimization*, Expert Systems with Applications, 38, 3353-3363.

Chiam S., Tan K., Mamum A.A. (2008), *Evolutionary Multi-objective Portfolio Optimization in Practical Context*, International Journal of Automation and Computing, 5(1), 67-80.

Deb K., Steuer R.E., Tewari R., Tewari R. (2011), *Bi-objective Portfolio Optimization Using a Customized Hybrid NSGA-II Procedure* [in:] *EMO 2011*, eds. R.H.C. Tekahashi et al., LNCS, Springer-Verlag, Berlin, Heidelberg, 6576, 358-373.

Ehrgott M. (2005), *Multicriteria Optimization*, Springer.

Elton E.J., Gruber M.J., Brown S.J., Goetzmann W.N. (2014), *Modern Portfolio Theory and Investment Analysis*, Wiley.

Gondzio J., Grothey A. (2007), *Solving Nonlinear Portfolio Optimization Problems with the Primal-Dual Interior Point Method*, European Journal of Operational Research, 181(3), 1019-1029.

Kaliszewski I. (2006), *Soft Computing for Complex Multiple Criteria Decision Making*, Springer.

Kaliszewski I., Miroforidis J. (2009), *Multiple Criteria Decision Making: Efficient Outcome Assessments with Evolutionary Optimization* [in:] Communications in Computer and Information Science Volume 35, Cutting-Edge Research Topics on Multiple Criteria Decision Making, 20th International Conference, MCDM 2009, Chengdu/Jiuzhaigou, China, June 21-26, 2009, Proceedings, Springer, 25-28.

Kaliszewski I., Miroforidis J. (2010), *Multiple Criteria Decision Making: From Exact to Heuristic Optimization* [in:] *Multiple Criteria Decision Making '09*, eds. T. Trzaskalik, T. Wachowicz, The University of Economics in Katowice, 113-120.

Kaliszewski I., Miroforidis J. (2012), *Real and Virtual Pareto Set Upper Approximations* [in:] *Multiple Criteria Decision Making '11*, eds. T. Trzaskalik, T. Wachowicz, The Publisher of University of Economics in Katowice, 121-131.

Kaliszewski I., Miroforidis J., Podkopaev D. (2012a), *Interactive Multiple Criteria Decision Making Based on Preference Driven Evolutionary Multiobjective Optimization with Controllable Accuracy*, European Journal of Operational Research, 216, 188-199.

Kaliszewski I., Miroforidis I., Podkopaev D. (2016), *Multiple Criteria Decision Making by Multiobjective Optimization - A Toolbox*, Springer.

Lo A.W., Petrov C., Wierzbicki M. (2003), *It's 11pm – Do You Know Where Your Liquidity Is? The Mean-variance-liquidity Frontier*, Journal of Investment Management, 1(1), 55-93.

Markowitz H.M. (1952), *Portfolio Selection*, Journal of Finance, 7, 77-91.

Markowitz H.M. (1991), *Portfolio Selection: Efficient Diversification of Investments*, Wiley, Monograph / Cowles Foundation for Research in Economics at Yale University, Wiley.

Miettinen K.M. (1999), *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers.

Miroforidis J. (2010), *Decision Making Aid for Operational Management of Department Stores with Multiple Criteria Optimization and Soft Computing*, PhD Thesis, Systems Research Institute, Warsaw.

Steuer R.E., Qi Y., Hirschberger M. (2011), *Comparative Issues in Large-scale Meanvariance Efficient Frontier Computation*, Decision Support Systems, 51(2), 250-255.

Steuer R.E., Qi Y., Hirschberger M. (2006), *Portfolio Optimization: New Capabilities and Future Methods*, Journal of Business Economics, 76(2), 199-220.

Wierzbicki A.P. (1999), *Reference Point Approaches*, Multicriteria Decision Making Advances in MCDM Models, Algorithms, Theory, and Applications, International Series in Operations Research and Management Science, Springer, 237-275.

MULTIPLE CRITERIA DECISION MAKING

Vol. 12 2017

Przemysław Juszczuk*
Lech Kruś**

SUPPORTING MULTICRITERIA FUZZY DECISIONS ON THE FOREX MARKET

DOI: 10.22367/mcdm.2017.12.05

Abstract

This paper deals with decisions made by a decision maker using technical analysis on the Forex market. For a number of currency pairs on the market the decision maker obtains buy or sell signals from transaction systems using technical analysis indicators. The signal is generated only when the assumed conditions are satisfied for a given indicator. The information characterizing every market situation and presented to the decision maker is binary: he either obtains the signal or does not.

In this paper a fuzzy multicriteria approach is proposed to extend and valuate information for the analysis of the market situation. The traditional approach with binary characterization of the market situations, referred to as a crisp approach, is replaced by a fuzzy approach, in which the strict conditions for which the crisp signal was generated are fuzzy. The efficiency of a given currency pair is estimated using values from the range $\langle 0, 1 \rangle$ and is defined by the membership function for each technical indicator. The values calculated for different indicators are treated as criteria. The efficiency of a given currency pair can be analyzed jointly for several indicators. The currency pairs are compared in the multicriteria space in which domination relations, describing preferences of the decision maker, are introduced. An algorithm is proposed which generates Pareto-optimal variants of currency pairs presented to the decision maker. The method proposed allows to extend the number of analyzed currency pairs, without significantly increasing the computation time.

* University of Economics in Katowice, Faculty of Informatics and Communication, Department of Knowledge Engineering, e-mail: przemyslaw.juszczuk@ue.katowice.pl.

** Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, e-mail: Lech.Krus@ibspan.waw.pl.

Keywords: foreign exchange, fuzzy sets, decision support.

1 Introduction

The Forex market is a global, decentralized currency market. It is regarded as the most liquid market in the world, and its daily turnover, even in the local currencies, often reaches trillions of dollars (McLeod, 2014). A currency pair is the basic instrument used on the Forex market; it can be regarded as the ratio of one currency to another. Thanks to four overlapping sessions, the Forex market has high accessibility even for retail clients. The ease of access to trading tools makes the Forex market very popular. Nowadays, such notions as technical analysis (which includes trend indicators, oscillators, Fibonacci levels, Pivot points and more), along with fundamental analysis, are crucial components of rule-based trading systems, in which easily understandable signals are used to open trading positions.

A rule-based trading system can be regarded as a set of rules related to technical analysis indicators or candle formations which are transformed into trading signals. Nowadays there is a growing tendency, in which these concepts are included in various complex systems based on neural networks (Yao et al., 2000; Lai et al., 2005), evolutionary prediction (Slany, 2009), evidence theory (Liu et al., 2009) and more. While fuzzy sets are also used in these methods (Kablan, 2009), many papers deal with multi-agent systems (Barbosa et al., 2008). Notions from AI systems (Yu et al., 2005) along with evolutionary computation (Hirabayashi et al., 2009) are intensively developed. Papers on fundamental analysis (like Nassirtoussi et al., 2015) included in the trading systems are less common that those including the technical analysis.

Nowadays there is no agreement about the efficiency of technical analysis. The same can be stated for rule-based trading systems. The number of papers dealing with the optimization of technical indicators remains high, while such papers as Cheol-Ho et al. (2007) indicate that these notions can be effective only in very specific situations, or are not effective at all. Article mentioned above is not an isolated work. Even in the case of High Frequency Trading systems one may observe that efficiency of automated trading systems decrease over time (Serbera et al., 2016). We propose to fill this gap by introducing a specially constructed system supporting the decision maker in the trading process. By a decision maker we understand a trader or any retail client with access to the Forex market. His/her goal is to achieve the highest possible profit from orders opened using signals generated from the set of trading rules. Nowadays, however, we can see a tendency for the number of potential instruments available to the decision maker on the Forex market to exceed a hundred. While the terms "decision maker" and "trader" in the present paper can be used simultaneously, we prefer to use decision analysis terms; therefore, in what follows we use the term "decision maker".

The approach introduced in this paper allows to initially estimate the set of currency pairs of potential interest for the decision maker. It should be clearly stated that our motivation is only to support the decision maker, and not to present an automatic trading system, so the main advantage of the proposed solution is that it assures the sovereignty of the decision maker (trader). The final decision whether to open a transaction for the given currency pair belongs to the decision maker. Thus, estimated set of variants can be regarded as a set of preliminary suggestions presented to the decision maker.

We propose an extension of the crisp approach currently used, where the signal is generated only in very specific market situations. In the crisp approach, the decision maker has a limited time to open the transaction when the signal has been generated. Thus the mechanism based on the binary values "signal / no signal" seems inefficient and for a large number of instruments often leads to the situation in which there is no single currency pair of potential interest to the decision maker.

In the fuzzy approach proposed here, there is a possibility to open a transaction in a predefined time interval related to the willingness of the decision maker to take the risk. This approach guarantees that the fuzzy approach is a generalization of the crisp approach, and the signals generated in the crisp approach are also included.

The outline of this paper is as follows. After the introduction, in Section 2, the notion of the crisp approach commonly used on the Forex market is briefly described. Next, the fuzzy approach along with the definitions of the membership functions are proposed. Section 3 includes a description of the proposed dominance-based algorithm generating non-dominated variants for the decision makers. Section 4 presents preliminary experiments conducted on real-world data. Finally, we present conclusions and suggest directions for future research.

2 Crisp and fuzzy systems

In the classical crisp approach, the rule-based trading system includes a predefined set of transaction rules involving technical indicators. Every indicator can be described by a set of rules which can be transformed into a binary activation function. A signal is generated only when the value of the function is equal to 1. In the fully automated-trading system a positive value of the function corresponds to opening the transaction, while in the crisp decision support system the information about the signal is derived in the system and presented to the decision maker.

We propose a fuzzy approach in which information about a market situation is transformed into a value of the membership function for each indicator. This value is calculated for every currency pair. Therefore, each currency pair in a given market situation at a time t is represented in the analysis as a variant with the vector of criteria related to particular indicators. To estimate the efficiency of this approach, we compare it with the classical crisp approach, where criteria for all indicators are constructed using the binary activation function. To accurately describe the proposed notion, we consider buy signals, but the same idea can be used for short sells.

To be more specific, we use two very popular technical indicators: the Relative Strength Index (RSI) and the Commodity Channel Index (CCI). The CCI indicator was originally proposed in the 1980s by Donald Lambert. The rules explaining the indicator are described in (www 1). A description of the RSI indicator can be found in Wilder (1978). These indicators are based on the so-called oversold and overbought levels and are frequently used to predict potential price changes. An example price chart with these indicators is shown in Figure 1. The overbought and oversold levels are in the upper and lower parts of the indicator windows. We have used the default parameters for the indicators with the overbought levels equal to 100 (for CCI), and 70 (for RSI). The oversold levels are equal to -100 and 30. In this particular example, the trading rule can be considered as the situation in which the indicator (CCI or RSI) crosses one of the levels defined above. If it crosses the oversold level upwards, the buy signal is generated. The sell signal is generated in the opposite case, when the overbought level is crossed downwards.

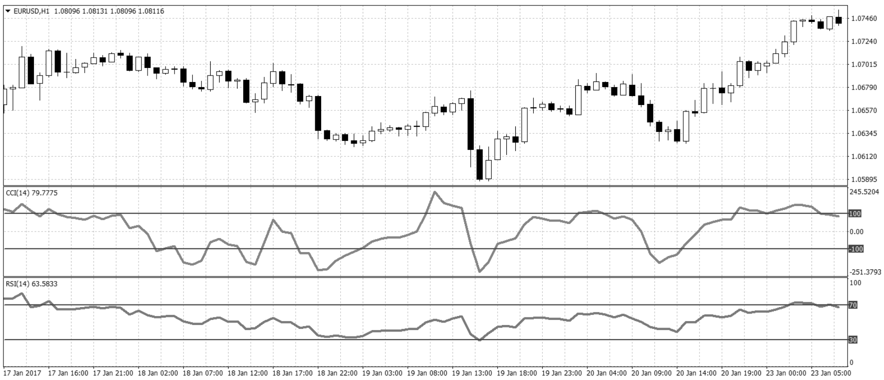


Figure 1. Example indicators and the price chart

The crisp signals for the RSI and CCI indicators are given by the formulas:

$$cond_{RSI} = true \text{ if } (RSI_n(t-1) < c_1) \wedge (RSI_n(t) > c_2), \tag{1}$$

$$cond_{CCI} = true \text{ if } (CCI_n(t-1) < c_1) \wedge (CCI_n(t) > c_2), \tag{2}$$

where $RSI_n(t-1)$ is the value of RSI at the time $t-1$ for the period n; $CCI_n(t-1)$ is the value of CCI at the time $t-1$ for the period n; c_1 and c_2 are constants related to their overbought and oversold levels.

We propose the fuzzy approach, in which the original signal taken from the crisp approach is still included. However, the adjacent values of the indicator can be also included by calculating the membership function:

$$\mu_{RSI}(c) = \begin{cases} \frac{RSI_n(t)}{30} \text{ if } (RSI_n(t) < 30), \\ 1 \text{ if } ((RSI_n(t-1) < 30) \wedge (RSI_n(t) > 30)) \\ \vee(RSI_n(t) = 31), \\ \frac{0.9}{RSI_n(t)-30} \cdot \alpha \text{ if } (RSI_n(t) > 31) \\ \wedge(RSI_n(t) < 50) \wedge (RSI_n(t-1) \leq 30), \\ 0 \text{ if } (RSI_n(t) > 50). \end{cases} \tag{3}$$

where α is a scalarizing factor in the range $\langle 0.5, 1.1\rangle$ and c is the currency pair for which the conditions on the right hand side of the equation are checked. The transaction system collects information from the market and calculates the values of the indicator at a given time t. Using the indicator the system checks the conditions and derives the value of membership function. The membership function for the CCI indicator is calculated as follows:

$$\mu_{CCI}(c) = \begin{cases} 0 \text{ if } (CCI_n(t) < CCI_{min}), \\ \frac{CCI_n(t)-CCI_{min}}{-CCI_{min}-100} \text{ if } (CCI_n(t) > CCI_{min}) \\ \wedge(CCI_n(t) < -100), \\ 1 \text{ if } (CCI_n(t-1) < -100) \wedge (CCI_n(t) > -100), \\ \frac{CCI_n(t)+50}{-50} \text{ if } (CCI_n(t) > -100) \wedge (CCI_n(t) < -50) \\ \wedge(CCI_n(t-1) > -100), \\ 0 \text{ if } (CCI_n(t) > -50). \end{cases} \tag{4}$$

where CCI_{min} is the minimal possible value of CCI and CCI_{max} is the maximal possible value of CCI. In the crisp case a signal can be observed only at a specific time tick – usually, when the indicator value is derived. In most cases the rules in the crisp approach use two adjacent values of the indicator. When the relation between these two values is satisfied (as in equation (1) or (2)), the signal for the decision maker is generated. In the fuzzy case the signal can be generated when the value of the membership function is higher than zero. Therefore the signal can be observed within a period longer than in the crisp approach and the decision maker has more time to make a decision.

3 A dominance-based algorithm

We will consider the buy signals. The sell signals can be treated in the same way. Let c be a currency pair valuated by a vector y of two criteria, $y = (y_1, y_2)$.

Variants of the vectors are analyzed in the criteria space $\mathbb{R}^2$. The criteria refer to the RSI and CCI indicators with the values of membership functions: $y_1 = \mu_{RSI}(c)$ and $y_2 = \mu_{CCI}(c)$ for a given currency pair c. The transaction system generates several such variants in a given time window. By a time window we understand a time needed to generate a new value on the price chart.

The decision maker, i.e. trader, tries to find a variant with the maximal values possible of all the criteria; therefore the following relations between variants are considered in $\mathbb{R}^2$ space:

Definition 3.1 *A variant y is at least as preferred as a variant z if each criterion of y is not worse than the respective criterion of z:*

$$y \succeq z \Leftrightarrow (y_1 \geq z_1) \wedge (y_2 \geq z_2). \tag{5}$$

Definition 3.2 *A variant y is more preferred (better) than a variant z if the following holds:*

$$y \succ z \Leftrightarrow (y \succeq z) \wedge \neg(z \succeq y). \tag{6}$$

Definition 3.3 *A variant y is incomparable with a variant z if:*

$$\neg(y \succeq z) \wedge \neg(z \succeq y). \tag{7}$$

The domination relation 6 defines a partial order in the space of criteria. We propose algorithm 1 for deriving non-dominated variants to be analyzed by the decision maker. The following notation is used in the algorithm.

- Y is the set of all variants considered in a given time window.
- $u = (1, 1)$ is assumed to be the aspiration point of the decision maker. If there exists a variant equal to the aspiration point, it should be considered as the only rational choice for the decision maker.
- x is the reservation point assumed by the decision maker. All variants dominated by this point are removed from further analysis. The reservation point x relates to the willingness of the decision maker to take a risk by extending the set of accepted variants as compared with the crisp approach. To be more specific, greater risk leads to the possibility of accepting potentially worse variants instead of delivering an empty set of variants to the decision maker.
- ND denotes the set of all non-dominated variants of potential interest to for the decision maker in a given time window.

- Y_- is the set of points removed from the analysis in the algorithm, initially the points dominated by x, i.e. $Y_- = (x + \mathbb{R}^2_- \setminus \{0\})$, where $\mathbb{R}^2_-$ is the negative cone. Moreover, it is the set of all points dominated by x and by the variants currently included in ND.

- Y_+ denotes the set of points accepted for further analysis in the algorithm, $Y_+ = Y \setminus Y_-$.

The algorithm, called the Dominance-based algorithm, allows for the generation of all non-dominated variants in the initial set Y_+ accepted for analysis on the basis of the reservation point x defined by the decision maker.

The steps of the algorithm can be divided into three phases. In the first phase the set Y is derived by calculating the criteria: membership functions for currency pairs in the assumed time window. At the same time, the sets Y_- and Y_+ are derived using reservation point x (lines 1–3). In the second phase (lines 4–5) variants equal to the aspiration point are looked for. If such a variant/variants exists, the algorithm is halted and the resulting set ND includes only these variants to be selected by the decision maker as his obvious rational choice. The third, and the most complex phase of the algorithm consists of lines 6–18. Three situations can occur: first if a selected variant y is included in Y_-, then it is removed from the analysis; second: if ND is empty, then the variant y is added to ND. The third situation occurs when ND is not empty. In this case variant y is compared with every variant from ND. The variants from ND dominated by y are removed from ND and y is added to ND. If the variant y is dominated by any variant of ND, the variant y is removed from the analysis. After each of these three situations Y_- is updated so that the area which it covers is expanded by the negative cone moved to y.

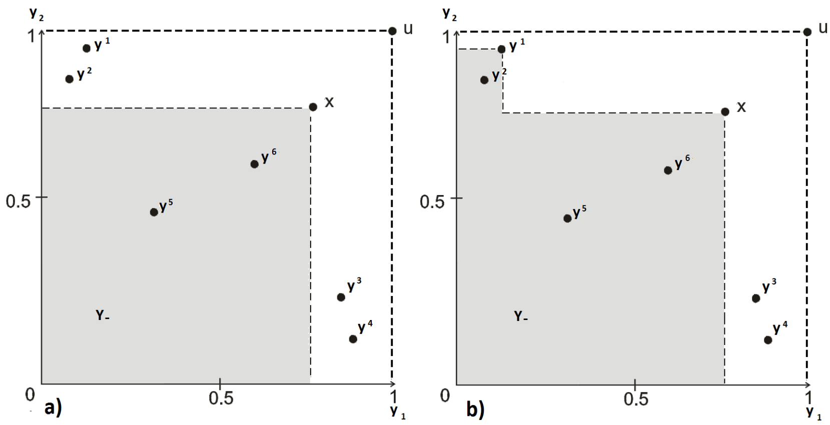


Figure 2. An ilustrative example

Algorithm 1: Dominance-based algorithm

begin
  Fix the aspiration point u, create the sets Y and $ND = \emptyset$
  The decision maker sets the point x defining the nonaccepted variants
  Generate sets Y_- and Y_+
  if *there exists* $y \in Y$ *such that* $y = u$ **then**
    $ND = \{y\}$ End of the algorithm
  for *each variant* y *in* Y_+ **do**
    if $y \in Y_-$ **then**
      Delete y from further analysis, i.e. $Y_+ = Y_+ \setminus \{y\}$
    else if $y \notin Y_- \wedge ND = \emptyset$ **then**
      Add y to ND and Update Y_- and $Y_+ = Y_+ \setminus \{y\}$.
    else
      for $z \in ND$ **do**
        if $y \succ z$ **then**
          Delete z from ND
        else if $z \succ y$ **then**
          Mark y as dominated, delete it from Y_+, and BREAK
      if y *is non-dominated* **then**
        Add y to ND
        Update $Y_- = Y_- \cup (y + \mathbb{R}^2_- \setminus \{0\})$
        Delete y from further analysis, i.e. $Y_+ = Y_+ \setminus \{y\}$
  if $Y_+ = \emptyset$ **then**
    end the algorithm

An illustrative example is presented in Figure 2. In part a) an initial simple situation is shown with a given reservation point x and a set Y consisting of six variants. Variants y^5 and y^6 are removed from further analysis because they are dominated by the reservation point x, as they belong to the initial set Y_- (marked in grey). An analysis of four remaining variants (y^1, y^2, y^3 and y^4) is illustrated in Figure 2b). At the beginning, ND is empty, and y^1 is outside the grey area, thus it is added to ND. The set Y_- is extended as follows:

$$Y_- = Y_- \cup (y^1 + \mathbb{R}^n_- \setminus \{0\}). \tag{8}$$

After the update of Y_-, variant y^2 is an element of Y_-, thus it is excluded from further analysis. Variants y^3 and y^4 are mutually incomparable and incomparable with y^1 (y^1 is already in ND). In this particular scenario both variants are added to ND. The grey area representing Y_- is expanded again. There are no more variants left, thus the algorithm halts. All non-dominated variants $\{y^1, y^3, y^4\}$ are in ND and can be presented to the decision maker.

4 Preliminary experiments

For the tests with real data we have selected 68 variants (currency pairs) from January 2017. We tested three time windows: 5 minutes (high frequency trading), 1 hour (intraday trading), 1 day (long-term trading). For every time window we have assumed three different positions of the reservation point in the criteria space: $x = (0.25, 0.25)$, $x = (0.5, 0.5)$ and $x = (0.75, 0.75)$. For every combination of these parameters (time window and reservation point position) we included 15 successive readings. The overall time of the experiments was equal to the length of the single time window multiplied by the number of readings. By a reading we understand a single situation on the price chart which is generated in the specific time window. To simplify: every new situation on the price chart corresponds to a new reading. The computation time for a single reading (including the generation of the non-dominated set ND) was approximately 2 seconds.

First of all, we derived variants which were not dominated by the reservation point. The numbers of such variants are presented in Table 1. They allow to estimate the potential number of variants (included in Y_+) which must be analyzed in detail. The table presents the cardinality of Y_+ for different time windows and different positions of the reservation point x. Lower values of the coordinates of the reservation point x indicate a higher willingness of the decision maker to take the risk. At the same time the cardinality of Y_+ is increased.

Table 1: The number of variants analyzed by the system with two indicators (RSI and CCI)

	M5			**H1**			**D1**		
x=	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**	**0.25**	**0.5**	**0.75**
Reading 1	15	14	9	15	8	6	18	13	9
Reading 2	15	14	11	16	14	6	14	8	7
Reading 3	22	16	12	22	19	13	20	13	11
Reading 4	25	23	19	29	25	19	19	17	12
Reading 5	33	28	23	27	24	23	23	19	11
Reading 6	26	19	16	12	11	9	30	22	16
Reading 7	21	18	14	18	13	6	29	26	19
Reading 8	20	11	10	24	20	17	37	31	20
Reading 9	17	14	7	21	16	10	33	28	27
Reading 10	22	15	12	21	13	8	25	22	17
Reading 11	15	10	3	25	20	17	23	20	17
Reading 12	11	11	9	27	21	19	20	16	13
Reading 13	19	15	10	19	15	12	25	18	11
Reading 14	22	20	15	21	16	11	33	22	22
Reading 15	17	13	9	27	22	18	20	15	15

After the preliminary selection of the variants included in the analysis, all the non-dominated variants were derived at the end of the given time window. The whole procedure was repeated for 15 successive readings. The results – the cardinality of the set ND for the time window of 5 minutes are presented in Table 2. The number of non-dominated variants derived for different positions of the reservation point decreases when the position of the point is moved in the direction of the aspiration point u. In the crisp approach the results are presented in the last column of Table 2. It should be noted that the rows with the same numbers for each column (Readings 2, 4 and 10) indicate the cases when variants equal to the aspiration point were found. These variants are presented to the decision maker as the only rational choices. Especially interesting are the cases in which the number of variants available for the decision makers is small. They can be observed in Readings 11 and 15, where in the first case the crisp approach generated no variant at all, while the fuzzy approach generated two variants for each value of x. In the second case, the crisp approach generated only one variant, while the number of variants derived from the fuzzy approach was: two for $x = 0.75$ and three for $x = 0.5$ and $x = 0.25$. In general, the set of solutions generated for the crisp case contains only the variants which were found in the corners of the criteria space, where the membership function for one of the indicators is equal to 1. The fuzzy approach generates all non-dominated variants generated in the crisp case. It also allows to extend the set of acceptable variants by the non-dominated variants for which a deviation from the aspiration point u is limited by the reservation point x. Namely, the non-dominated variants belong to the set $(u + \mathbb{R}^n_-) \setminus (x + \mathbb{R}^n_- \setminus \{0\})$.

Table 2: The number of non-dominated variants generated in the fuzzy case as compared with the number of variants generated in the crisp approach – – 5 minutes time window

	M5			
x =	**0.25**	**0.5**	**0.75**	**Crisp**
Reading 1	7	7	5	2
Reading 2	1	1	1	1
Reading 3	12	9	9	5
Reading 4	3	3	3	3
Reading 5	18	16	14	9
Reading 6	9	8	7	3
Reading 7	8	8	8	6
Reading 8	8	7	5	3
Reading 9	5	5	3	2
Reading 10	1	1	1	1
Reading 11	2	2	2	0
Reading 12	7	7	7	4
Reading 13	10	8	7	5
Reading 14	10	9	8	5
Reading 15	3	3	2	1

In Table 3 we present similar results for the cardinality of ND for the time window of 1 hour. Once again, the comparative results for the crisp approach are given in the last column. In Readings 2, 7 and 14 once again the number of variants derived in the crisp approach was very small, while the application of the fuzzy method increased the number of non-dominated variants derived for the decision maker. A shortcoming of the system can be observed in the readings 5, 12 and 13, where the advantage of the fuzzy approach is not visible due to a large number of variants derived for the decision maker from the crisp approach. In such situations an analysis based on a greater number of indicators should be made. If the fuzzy approach generates a large number of non-dominated variants, an appropriate ranking method should be applied.

Finally, the results for the longest time window considered on the Forex market as the long-term trading (1 day time window) which covered approximately three weeks from January 2017 are presented in Table 4. Once again, the most useful information for the decision maker is generated for Reading 2, where the crisp approach resulted in one variant only, while in the fuzzy case at least three variants have been generated.

Table 3: The number of non-dominated variants generated in the fuzzy case approach as compared to the number of variants generated in the crisp approach – 1 hour time window

	H1			
x =	**0.25**	**0.5**	**0.75**	**Crisp**
Reading 1	6	5	5	3
Reading 2	3	3	2	1
Reading 3	6	6	5	3
Reading 4	2	2	2	2
Reading 5	20	20	19	17
Reading 6	8	7	6	4
Reading 7	5	5	4	2
Reading 8	14	10	10	7
Reading 9	9	7	6	4
Reading 10	5	5	5	2
Reading 11	9	8	7	4
Reading 12	16	14	12	10
Reading 13	15	13	11	10
Reading 14	7	7	5	3
Reading 15	1	1	1	1

Table 4: The number of non-dominated variants generated in the fuzzy case approach as compared to the number of variants generated in the crisp approach – 1 day time window

	D1			
x =	**0.25**	**0.5**	**0.75**	**Crisp**
Reading 1	2	2	2	2
Reading 2	4	3	3	1
Reading 3	10	8	7	5
Reading 4	9	9	7	5
Reading 5	7	7	5	3
Reading 6	9	8	4	2
Reading 7	13	12	12	5
Reading 8	9	8	8	4
Reading 9	6	6	6	6
Reading 10	10	10	9	6
Reading 11	1	1	1	1
Reading 12	11	10	10	5
Reading 13	2	2	2	2
Reading 14	9	9	8	3
Reading 15	11	10	10	7

5 Conclusions

In this paper we have presented an extension of the classical crisp approach used in rule-based trading systems on the Forex market. The suggested approach provides an opportunity to extend the set of variants of possible interest for the decision maker. The fuzzy notion is presented as a generalization of the notions commonly used in rule-based trading systems. Along with the implementation of fuzzy membership functions, an algorithm generating the set of non-dominated solutions has been presented. The algorithm is especially useful when the traditional crisp approach generates no signals at all, but the fuzzy approach provides variants with the membership function close to 1. The notion of a reservation point is related to the risk aversion of the decision maker.

The proposed approach assures the full sovereignty of the decision maker. He decides how far he wants to extend the set of variants analyzed by the system in comparison to the crisp approach. He obtains the generated non-dominated variants. The decision maker decides which variant he will use to make a position.

The proposed approach is flexible. It can be used with various time windows and even with a different set of instruments. The position of the reservation point in the criteria space can be changed for every new reading. The approach, presented here for two indicators, can be easily extended to handle a greater number of them. A greater number of indicators included in the system should significantly reduce the set of non-dominated variants. Especially interesting can be such indicators as the moving average (Holt, 2009), money flow index, Ichimoku and others (Patel, 2010).

At the preliminary stage of our experiments we assumed that there are no complex dependencies between the two indicators. However, in general this is not strictly true for a greater number of different indicators. There are indicators which should be analyzed jointly under additional assumptions. Complex dependencies and complex transaction systems will be introduced in the proposed method. Finally, generating a large and difficult set of non-dominated variants naturally forces a ranking of the variants which could greatly improve the final analysis performed by the decision maker. An appropriate ranking method will be included in the system. Different ranking approaches are discussed, including ideas based on the notion of a concession line (Juszczuk et al., 2016), the dominance-based rough set approach (Greco et al., 2002), or the bipolar method (Konarzewska-Gubała, 1989).

As mentioned before, the presented approach provides an initial simple version of the system, which can be expanded in many ways. Among the increasing number of papers dealing with more complex systems based on the technical analysis indicators, two distinct extensions seem especially interesting. Both are related to social phenomena, which could be used in the systems. The first extension assumes the introduction of a fundamental analysis translated into easily understandable numeric values of the fundamental indicators. The second one is strictly related to social trading regarded as a mechanism for collaborative trading on the market. The effectiveness of social trading and systems based on social networks (such as twitter) is particularly difficult to estimate. On the other hand, it is possible to estimate the activity of traders using such notions as gamification. These concepts will therefore be discussed in future papers.

References

Barbosa R.P., Belo O. (2008), *Autonomous Forex Trading Agents*, Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects, Lecture Notes in Computer Science, 5077, Springer.

Cheol-Ho P., Irwin S.H. (2007), *What Do We Know about the Profitability of Technical Analysis?* Journal of Economic Surveys, 21(4), 786-826.

Greco S., Matarazzo B., Słowiński R. (2002), *Multicriteria Classification by Dominance-based Rough Set Approach*, Handbook of Data Mining and Knowledge Discovery, Oxford University Press, New York.

Hirabayashi A., Aranha C., Iba H. (2009), *Optimization of the Trading Rule in Foreign Exchange Using Genetic Algorithm*, Proceedings of the 11th Annual conference on Genetic and evolutionary computation GECCO '09, 1529-1536.

Holt C.C. (2004), *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*, International Journal of Forecasting, 20(1), 5-10.

Juszczuk P., Kaliszewski I., Podkopaev D., Hsu-Shih S. (2016), *Market Collective Wisdom Discovery for Portfolio Investments*, International Journal of Information and Management Sciences, 27, 87-102.

Kablan A. (2009), *Adaptive Neuro Fuzzy Inference Systems for High Frequency Financial Trading and Forecasting*, Advanced Engineering Computing and Applications in Sciences, ADVCOMP '09.

Konarzewska-Gubała E. (1989), *BIPOLAR: Multiple Criteria Decision Aid Using Bipolar Reference System, LAMSADE*, Cashier et Documents, 56, Paris.

Lai K.K., Yu L., Wang S. (2005), *A Neural Network and Web-Based Decision Support System for Forex Forecasting and Trading*, Data Mining and Knowledge Management, Lecture Notes in Computer Science, 3327, Springer.

Liu Z., Xiao D. (2009), *An Automated Trading System with Multi-indicator Fusion Based on D-S Evidence Theory in Forex Market*, Fuzzy Systems and Knowledge Discovery, FSKD'09.

McLeod G. (2014), *Forex Market Size: A Traders Advantage*, dailyFx, January 24, https://www.dailyfx.com/forex/education/trading (access: 25.04.2017).

Nassirtoussi A.K., Aghabozorgi S., Wah T.Y., Ling Ngo D.C. (2015), *Text Mining of News-headlines for FOREX Market Prediction: A Multi-layer Dimension Reduction Algorithm with Semantics and Sentiment*, Expert Systems with Applications, 42(1), 306-324.

Patel M. (2010), *Trading with Ichimoku Clouds: The Essential Guide to Ichimoku Kinko Hyo Technical Analysis*, Wiley & Sons.

Serbera J.P., Paumard P. (2016), *The Fall of High-frequency Trading: A Survey of Competition and Profits*, Research in International Business and Finance, 36, 271-287.

Slany K. (2009), *Towards the Automatic Evolutionary Prediction of the FOREX Market Behaviour*, Adaptive and Intelligent Systems, ICAIS '09.

Wilder W. (1978), *New Concepts in Technical Trading Systems*, Trend Research.

Yao J., Tan Ch.L. (2000), *A Case Study on Using Neural Networks to Perform Technical Forecasting of Forex*, Neurocomputing, 34(1-4), 79-98.

Yu L., Lai K.K., Wang S. (2005), *Designing a Hybrid AI System as a Forex Trading Decision Support Tool*, Tools with Artificial Intelligence, ICTAI '05.

(www 1) Investopedia, http://www.investopedia.com/articles/trading/05/041805.asp (access: 10.04.2017).

MULTIPLE CRITERIA DECISION MAKING

Vol. 12 2017

Bogumił Kamiński*
Michał Jakubczyk**

COMPARING THE CRISP AND FUZZY APPROACHES TO MODELLING PREFERENCES TOWARDS HEALTH STATES

DOI: 10.22367/mcdm.2017.12.06

Abstract

Understanding societal preferences towards health is vital in public decisions on financing health technologies. Thought experiments in which respondents choose between health states are used to understand the importance of individual criteria. Competing models of preference structure can be compared by their ability to explain empirical observations. One of the key challenges when constructing such models is that they have to aggregate preferences defined in multiple-criteria space. In the present paper, we test whether treating the impact of health worsening (defined using EQ-5D-5L descriptive system, i.e. decomposing health status in five criteria) as a fuzzy concept can improve the model fit. To test if fuzzy approach to multiple-criteria preferences aggregation is valid, we compare a standard, crisp model (SM) with two models using fuzzy sets (JKL, previously proposed in the literature; and FMN introduced here). We find FMN better than SM, and SM better than JKL. Anxiety/depression and pain/discomfort seem to weigh most in preferences. According to FMN, self-care and usual activities are associated with largest imprecision in preferences. The respondents are susceptible to framing effects when time unit is changed: e.g. measuring the duration in days results in short intervals mattering more than when expressed as weeks. We conclude that the fuzzy-based framework is promising, but requires careful work on the exact specification.

* SGH Warsaw School of Economics, Decision Analysis and Support Unit, Warsaw, Poland, e-mail:bogumil.kaminski@sgh.waw.pl.

** SGH Warsaw School of Economics, Decision Analysis and Support Unit, Warsaw, Poland, e-mail:michal.jakubczyk@sgh.waw.pl.

Keywords: fuzzy modelling, discrete choice experiment, health-related quality of life, preference elicitation.

1 Introduction

Understanding people's preferences towards health is important; it can serve in health technology assessment (HTA) to evaluate benefits of increasing life expectancy or improving health-related quality of life (HRQoL). Hence, the elicitation of preferences serves a prescriptive purpose: to suggest a course of action to be taken, typically, by the public regulator on behalf of the society.

There are two steps in the preference-elicitation process. First, the mathematical representation of the preferences is constructed; then, its parameters are estimated based on empirical data. Regarding the former, typically a quality-adjusted life years (QALY) model is used in HTA (Weinstein et al., 2009): a health state Q is assigned a number, $u(Q)$, with the interpretation that $T \times u(Q)$ denotes the von Neumann-Morgenstern utility of spending T years in Q, where $u(\text{dead}) = 0$ and $u(\text{full health}) = 1$ (see Bleichrodt et al., 1997; Miyamoto et al., 1998). The key challenge is that Q is usually evaluated using multiple criteria, and u is a function aggregating them into a single instrument that can be used operationally. In this text, we use the EQ-5L-5D system to define HRQoL, which implies a five-attribute description of Q.

Regarding the parameter estimation, usually either a time trade-off (TTO) or a discrete choice experiment (DCE) is used to collect data on preferences. In TTO, we attempt to determine the time T, such that T years in full health is equivalent to 10 years in Q; in DCE, the respondent faces a series of pairwise comparisons between two states, Q_1 and Q_2, lasting for T_1 and T_2 (in *DCE with duration*), respectively (immediate death may also be used). The results of TTO or DCE, combined with the QALY model assumptions (and the resulting econometric models), can serve to assign utilities to health states.

However popular in applied HTA the QALY model is, its founding assumptions are often criticized (e.g. Attema et al., 2010; Pettitt et al., 2016; Beresniak et al., 2015). Among various lines of critique, Jakubczyk and Kamiński (2017) and Jakubczyk (2015) suggested that fuzzy sets (a concept introduced by Zadeh, 1965) can be used to define preferences towards health states (in the context of health vs money trade-offs); the approach is motivated by the observation that a lack of market experience can lead to an inherent imprecision in preferences. In the present paper, we aim to compare the standard (crisp) approach with a fuzzy-based one. Even though there is no descriptive motive in the health preference research – we do not strive to predict somebody's choices (as people rarely actually choose between health states) – the model fit seems a natural way to evaluate the credibility of the elicited values (Jakubczyk et al., 2017).

We compare three approaches: 1) a standard QALY-model-based, crisp model, as a benchmark; 2) a fuzzy-based approach proposed by Jakubczyk et al. (2017) (JKL, henceforth); 3) an alternative fuzzy-based specification, developed in the present paper (FMN, henceforth). We use the model fit as a basic measure of model quality (minus log likelihood), but also discuss the face validity of estimated parameters.

In the next section, we first present more details on how health states are defined and the specifications of all three approaches. In section 3, we introduce the dataset and the numerical approach used to estimate the parameters of the models. In section 4, we present the results, compare the models with respect to the insight on the impact of health on utility, and comment on the predictive validity of the approaches. Finally, we discuss our findings.

2 Fuzzy modelling of preferences towards health states

2.1 Benchmark, crisp model

Health states are often described with the EQ-5D-3L (or 5L) descriptive system (Brooks et al., 1996; Herdman et al., 2011), i.e. using five dimensions (or *criteria* in decision modelling parlance): mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). In each dimension, health can be at one of three (in 3L) or five (5L) levels, denoting no problems (level 1) or more and more severe problems (consecutive levels). In such a descriptive system, a health state is denoted by five consecutive digits; in particular, 11111 denotes full health (FH), and 55555 (we focus on 5L case henceforth) denotes the worst (in the descriptive system considered) possible health state[1].

There are 3125 health states in the EQ-5D-5L descriptive system, making it virtually impossible to elicit the utility for all of them. For this reason, a model is fitted to the data collected for a subset of states, and then the utilities of all the states can be approximated via extrapolation. Typically, the utility of a health state Q is calculated relative to the utility $u(\mathrm{FH}) = 1$, in the form:

$$u(Q) = 1 - \sum_{i=1}^{5} \sum_{j=2}^{5} \alpha_{i,j} d_{i,j}(Q), \tag{1}$$

where

[1] This notation is in standard use in the literature (and, e.g., not a vector-like $(5, 5, 5, 5, 5)$); hence, we use it here.

- i indexes the dimensions,
- j indexes the levels (no disutility for level 1; hence, omitted in the formula above),
- $d_{i,j}$ is a dummy denoting whether dimension i is at level j,
- parameters $\alpha_{i,j}$ represent the preference structure (again, no disutility at level 1).

We expect that $\alpha_{i,j}$ is positive and increasing with j. Often a constant term, α_0 is added (if Q differs from FH, i.e. if at least one $d_{i,j} = 1$). Because JKL did not use it in their specification and because α_0 is difficult to interpret, we omit it in the basic benchmark specification here[2]. Nonetheless, we also present the version with the constant term, differing little in terms of the model fit.

When the above model is estimated based on TTO data, an error term, ε, is added to eq. (1) (otherwise, no set of parameters α could fit the observed data). In DCE, when two health states, Q_A and Q_B, considered for T_A and T_B years, respectively, are compared, it is often assumed that the probability of Q_A being selected is given by an exponential version of the Bradley-Terry approach (e.g. Bansback et al., 2012):

$$P(Q_A, T_A, Q_B, T_B) = \frac{\exp\left(u(Q_A) \times T_A\right)}{\exp\left(u(Q_A) \times T_A\right) + \exp\left(u(Q_B) \times T_B\right)}. \tag{2}$$

In the benchmark model, we use eq. (2) to estimate the parameters of eq. (1).

Notice that equation (2) allows the utility to be negative (which is not a problem thanks to the exponential function), and indeed some health states are perceived as worse than dead. Immediate death is equivalent to 0 utility (i.e. is equivalent to any state with duration zero). Jakubczyk et al. (2017) point out that in the above formula there is always a positive probability of any state being selected, however worse it is (even dominated) than the other one.

In equation (2), the rescaling of T_A and T_B (i.e. multiplying both by the same positive constant) changes the probability. Therefore, whether the time is expressed as years (usually the case) or months, weeks, days, should also impact the result. At the same time, presenting the choice in the specific time unit might frame the problem differently (e.g. the subjective perception of *six months* might differ from *half a year*). In our dataset, we use various time units (days, weeks, months, and years). Hence, we introduce three parameters,

[2] The constant term was interpreted in terms of dimensions complementarity by Jakubczyk (2009); it could be also interpreted to reflect the fact that the state 11111 might also include some minor health problems.

τ_1, τ_2, and τ_3, scaling the duration when days, weeks, or months are used, respectively. For example, if time is expressed in days in the task, then the formula for probability becomes:

$$P(Q_A, T_A, Q_B, T_B) = \frac{\exp\left(u(Q_A) \times T_A \times \tau_1\right)}{\exp\left(u(Q_A) \times T_A \times \tau_1\right) + \exp\left(u(Q_B) \times T_B \times \tau_1\right)}. \quad (3)$$

The parameters to be estimated are α (20) and τ (3). We expect τ_i to be < 1 and increasing with i. If there is no framing effect of a time unit, we should get $\tau_1 = 1/365$, $\tau_2 = 1/52$, and $\tau_3 = 1/12$.

2.2 JKL fuzzy model

We only briefly reintroduce the JKL model (Jakubczyk et al., 2017), and the reader is encouraged to see the original publication for details. JKL suggested to treat the utility of being in state Q for T years, $u(Q, T)$, as a fuzzy set. For simplicity, they used a piecewise linear membership function, $\mu_{u(Q,T)}(x)$; $\mu_{u(Q,T)}(x) = 1$ for low values ($x \le L(Q) \times T$), $\mu_{u(Q,T)}(x) = 0$ for high values ($x \ge H(Q) \times T$), where $L(Q)$ and $H(Q)$ are parameters characterizing health state Q. Then, $\mu_{u(Q,T)}(x)$ is linearly decreasing between $L(Q) \times T$ and $H(Q) \times$ $\times$ T (or jumping discontinuously, if $L(Q) \times T = H(Q) \times T$). JKL interpret $\mu_{u(Q,T)}(x)$ as the conviction that being in Q for T years gives the utility of at least x.

Analogously to eq. (1), $L(Q)$ and $H(Q)$ are given as linear combinations of dummies for dimensions and levels defining Q:

$$L(Q) = 1 - \sum_{i=1}^{5} \sum_{j=2}^{5} h_{i,j} d_{i,j}(Q), \quad (4)$$

$$H(Q) = 1 - \sum_{i=1}^{5} \sum_{j=2}^{5} l_{i,j} d_{i,j}(Q). \quad (5)$$

Parameters l and h define the range of disutility for a given dimension/level. The larger they are, the bigger the impact of a given health worsening on utility. The more they differ, the larger the imprecision in the perception of disutility. Because of the subtraction, parameters h are used to define $L(Q)$, and parameters l are used to define $H(Q)$.

In the JKL model, two health states are compared in the following way (e.g. in a DCE experiment). The advantage of (Q_1, T_1) over (Q_2, T_2), namely $\delta_{(Q_1,T_1),(Q_2,T_2)}$, is given as:

$$\sup_{x \in \mathbb{R}} \left(\mu_{u(Q_1,T_1)}(x) - \mu_{u(Q_2,T_2)}(x) \right),$$

and the advantage the other way round is defined analogously. The parameters δ must be in the $[0, 1]$ range. It can happen that both δs are positive. Then, the net advantage of advantage of (Q_1, T_1) over (Q_2, T_2), $\Delta_{(Q_1,T_1),(Q_2,T_2)}$, is given as $\delta_{(Q_1,T_1),(Q_2,T_2)} - \delta_{(Q_2,T_2),(Q_1,T_1)}$, and the resulting $\Delta \in [-1, 1]$.

The probability of (Q_1, T_1) being chosen instead of (Q_2, T_2) is given as a function of the net advantage: $P(\Delta) = \frac{(\Delta+1)^\rho}{2}$, for $\Delta \leq 0$, with a non-negative parameter ρ to be estimated (and for $\Delta > 0$, the probability is calculated using the assumption that $P(\Delta) + P(-\Delta) = 1$). The value $\rho = 1$ leads to Δ being transformed linearly into probability, $\rho < 1$ results in probability remaining at around 50% for many values of Δ, and $\rho > 1$ results in a probability being sensitive to Δ values differing even slightly from zero.

Let us notice the following features of JKL's model. First, multiplying T_1 and T_2 by the same strictly positive number does not change the preferences (parameters δ, Δ, and $P(\cdot)$). Hence, no counterparts of τ are needed (as long as the same time unit is used in both states compared). Second, for large enough differences between (Q_1, T_1) and (Q_2, T_2), the probability of one being chosen is equal to 1 (not only approaches 1). Third, this model compares the two health profiles in the conviction space (i.e. the values of membership functions are compared), rather than in the utility space. This last property motivates trying another fuzzy-based approach, presented in the next subsection (in which the first two properties do not hold; hence, the model is more flexible).

2.3 Fuzzy model – a new specification (FMN)

Again, the utility of living in Q for T years, $u(Q, T)$, is defined by two numbers, $L(Q) \times T$ and $H(Q) \times T$. The values $L(Q)$ and $H(Q)$ are defined as in eq. (4). In the present specification, though, the membership function, $\mu_{u(Q,T)}(x)$ is equal to 1 for $L(Q) \times T \leq x \leq H(Q) \times T$, and 0 otherwise. The interpretation is that the decision maker agrees fully that T years in state Q may correspond to the utility of x, or – putting it differently – cannot rule out x as the utility of (Q, T) and totally rules out any value below $L(Q) \times T$ or above $H(Q) \times T$.

We then define the advantage of one profile, (Q_1, T_1), over another, (Q_2, T_2), by comparing the middles of 1-cuts of $u(Q, T)$:

$$\Delta_{(Q_1,T_1),(Q_2,T_2)} = \left(T_2 \frac{H(Q_2) + L(Q_2)}{2} - T_1 \frac{H(Q_1) + L(Q_1)}{2} \right) \times \tau_i. \qquad (6)$$

The parameter τ_i is the scaling factor, and i changes with the time unit used to measure T_1 and T_2 (the same time unit assumed): $i = 1$ for days, $i = 2$ for weeks, $i = 3$ for months, and $i = 4$ for years. We normalize $\tau_4 = 1$, and estimate τ_1, τ_2, and τ_3.

We define the ancillary score, π, (subscripts $(Q_1, T_1), (Q_2, T_2)$ suppressed; cf. eq. (2)),

$$\pi = \frac{1}{1 + \exp(-\Delta)}. \tag{7}$$

In this way, we transform Δ in the $\pi \in [0, 1]$ interval, to facilitate interpreting the advantage in terms of probabilities.

In this model, we try to account for the fact that a larger difference between $L(Q)$ and $H(Q)$ denotes a larger imprecision in how the utility of Q is perceived. We want to include the possibility that larger imprecision may dilute the preferences, i.e. shift the probability of one alternative being chosen towards 50%. Specifically, we take:

$$\Theta_{(Q_1,T_1),(Q_2,T_2)} = \left(T_2 \frac{H(Q_2) - L(Q_2)}{2} + T_1 \frac{H(Q_1) - L(Q_1)}{2} \right) \times \tau_i. \tag{8}$$

In the present specification, we use the same vector $[\tau_1, \tau_2, \tau_3]$ when calculating Δ and θ, assuming the time unit is perceived identically in both aspects.

Finally, we define the resulting probability of (Q_1, T_1) being chosen as:

$$P = \frac{\pi - 0.5}{1 + \omega\Theta} + 0.5, \tag{9}$$

where ω is a parameter to be estimated. For $\omega = 0$ there is no impact of imprecision on preferences (and the model can be reduced to a crisp version with $L(Q) = H(Q)$ for all the states).

Summing up, in the FMN specification, we have to estimate parameters h and l (40), ω (one parameter), and τ (3).

3 Methods

3.1 Dataset

We used the data from the DCE predictive competition organised by the International Academy for Health Preference Research (IAHPR), sponsored by The EuroQol Group, available for general public, and described on the IAHPR website (`http://iahpr.org/eq-dce-competition/`, as of 16 Nov, 2016). More details can be found in Jakubczyk et al. (2017).

There were responses from 4074 US respondents, each choosing between two health states in 20 pairwise comparisons. Each health state was described with the EQ-5D-5L descriptive system, and the duration was given (four time units were used: days, weeks, months, and years; the same unit for both states in each pair).

In the modelling, only the aggregated proportion of a given answer is used, not the entire individual answer.

3.2 The approach to estimation

In the data set, there were 1560 different combinations of compared states and times. For each combination c_i, the number of observations, n_i, and the number of choices of option 1, k_i, were recorded.

To estimate the parameters of the model (in all three specifications), we employed the maximum likelihood estimation:

$$\max \quad \mathcal{L} = \sum_{i=1}^{1560} k_i \ln(\Pr(c_i)) + (n_i - k_i) \ln(1 - \Pr(c_i)), \tag{10}$$

by changing the parameters used in a given specification. Although the three models have significantly different specifications, using the same objective function allows us to analyse their fit by simply comparing the estimated $-\log(\mathcal{L})$ (the lower the better).

In the estimation for level 1, we assumed that the preference is crisp and equal to 0 for every dimension. When estimating the parameters of the crisp specification, we imposed the constraints that $\alpha_{i,j}$ are positive and non decreasing in j for every dimension i. When estimating the parameters of the fuzzy specifications we imposed the following constrains: (a) $l_{i,j}$ and $h_{i,j}$ are positive, (b) for every dimension i and level j, $h_{i,j} \geq l_{i,j}$, (c) the mean of $l_{i,j}$ and $h_{i,j}$ is non-decreasing in j for every dimension i. They reflect the assumptions made in the crisp model and only add the fuzzy set consistency restriction (b).

The problem specified in eq. (10), subject to the above constraints, was solved using the Nelder-Mead optimization in both cases: crisp and the new fuzzy FMN specification. The constraints were imposed by adding constraint violation penalty. In the latter, in order to ensure an effective estimation of the objective function, we replaced functions $l_{i,j}$ and $h_{i,j}$ by $(h_{i,j} + l_{i,j})/2$ and $(h_{i,j} - l_{i,j})/2$, as the latter form was simpler to impose constraints on and, in consequence, it provided a higher stability of the results. In the JKL specification, eq. (10) is non-differentiable due to the way we compare fuzzy sets. Therefore, Nelder-Mead or other standard optimization routines failed to consistently converge. To overcome these problems, we employed simulated annealing optimization in this case.

For the Nelder-Mead optimization, we used the implementation that follows Gao et al. (2012) and uses default parameters and stopping criteria, except for the number of iterations (we did not impose any restriction on the number of iterations of the optimization procedure). It was implemented in the Julia language (Bezanson et al., 2017) using the Optim package (White et al., 2017). The specification of the crisp model is effectively a logistic regression and thus it has a single local (and thus global) optimium (Menard,2002).

The FMN model, with $(h_{i,j} + l_{i,j})/2$ and $(h_{i,j} - l_{i,j})/2$ as decision variables, is a logistic regression with an additional monotonic transformation given by (9). We do not have a proof that this transformation has a single minimum, but intuitively the properties of the objective function should not be significantly different. To verify the stability of the solution, we ran the optimization 1000 times, each time starting from a new point sampled uniformly from the admissible region; the optimization converged to approximately the same solution.

For the estimation of the JKL model parameters, we found that this model has multiple local minima. Therefore, we developed a custom algorithm based on simulated annealing (Du and Swamy, 2016). The exact procedure was the following. We started with an admissible point sampled uniformly. Then, 10,000 steps of simulation annealing were performed with the application of Gaussian perturbations to all parameters (inadmissible perturbations were rejected). In the second stage, to perform a search near the optimum, we performed a random local search in which in each step we perturbed only one parameter and accepted the new solution only if it improved the solution. The second step was halted when for a batch of 1000 iterations the improvement of the objective function was less than 10^{-8} (approximately the square root of the precision of IEEE 754 floating point around 1.0). To verify that we do not end in a local minimum, we applied the multi-start (Martí, 2003) approach – the procedure was run 1000 times starting from a different random point. We report the best solution found. As with any heuristic approach, this is only an approximation of the optimal solution. However, it should be noted that better properties of the optimized objective function are another reason for preferring the FMN fuzzy approach presented in this paper over the JKL model.

4 Results

4.1 Crisp model

The results for the benchmark model are presented in Table 1, in the specification without and with the constant term α_0. In both approaches, the AD dimension was found to cause the greatest disutility (when at level 5) followed by PD; on the other hand, the UA dimension causes the least disutility (looking at level 5 only).

The estimated values of τ show that durations measured using various time units are not simply algebraically recalculated, e.g., into years. For example, $\tau_1 = 0.123 > 1/365$, showing that the relative importance of one day, when duration is measured in days, is greater than if it was measured in years.

Notice that even $\tau_2 > \tau_3$, suggesting that one week has greater weight than one month, but that may be due to estimation imprecision.

The measure of fit, that is, negative log of likelihood, has no direct interpretation (it will be compared to the ones obtained for other models).

The values presented in the table can be used to calculate the utility for all 3125 states in the EQ-5D-5L descriptive system. For example, $u(55555) = -0.668$ (for the specification without α_0).

Table 1: Crisp model, without and with the free parameter. Measure of fit equal to 52538 and 52410, respectively. Dimensions: MO = mobility, SC = self-care, UA = usual activities, PD = pain/discomfort, AD = anxiety/depression

Dimension/level	**Description**	**No α_0**	**With α_0**
Constant		—	0.128
MO2	slight problems in walking about	0.067	0.038
MO3	moderate problems ...	0.094	0.078
MO4	severe problems ...	0.241	0.210
MO5	unable to walk about	0.318	0.290
SC2	slight problems washing or dressing	0.039	0.025
SC3	moderate problems ...	0.069	0.063
SC4	severe problems ...	0.215	0.200
SC5	unable to wash or dress myself	0.319	0.297
UA2	slight problems doing usual activities	0.111	0.065
UA3	moderate problems ...	0.135	0.094
UA4	severe problems ...	0.263	0.227
UA5	unable to do usual activities	0.263	0.227
PD2	slight pain or discomfort	0.076	0.048
PD3	moderate ...	0.122	0.093
PD4	severe ...	0.358	0.328
PD5	extreme ...	0.358	0.328
AD2	slightly anxious or depressed	0.120	0.080
AD3	moderately ...	0.221	0.196
AD4	severely ...	0.410	0.367
AD5	extremely ...	0.410	0.367
T1	day	0.123	0.129
T2	week	0.385	0.400
T3	month	0.368	0.389

4.2 JKL model

In Table 2, we present the estimation results for the JKL model, as in the original publication: Jakubczyk et al. (2017). When focusing on level 5, the PD and AD dimensions were found to cause the greatest disutility. Also, AD is associated with largest imprecision of preferences: the difference between $h_{5,5}$ and $l_{5,5}$ amounts to almost 0.6, nearly two thirds of the difference in utility between dead and full health.

In JKL specification, the utility of the worst state, $u(55555)$, is a wide interval: $[-2.02; -0.07]$.

Importantly for the present paper, the fit decreases significantly when compared with the crisp approach, even though the number of parameters has doubled. This finding motivates trying another fuzzy approach, as specified in subsection 2.3, whose results are presented subsequently.

Table 2: Fuzzy model, dimensions as in Table 1, $\rho = 0.989$. Measure of fit equals 60971

Dimension	**Parameter**	**Level**			
		2	**3**	**4**	**5**
MO	l	0.034	0.034	0.200	0.320
	h	0.215	0.215	0.500	0.601
SC	l	0.000	0.026	0.116	0.208
	h	0.186	0.278	0.388	0.530
UA	l	0.018	0.018	0.138	0.138
	h	0.138	0.206	0.355	0.389
PD	l	0.000	0.071	0.210	0.266
	h	0.296	0.296	0.546	0.771
AD	l	0.031	0.091	0.091	0.138
	h	0.120	0.242	0.701	0.727

4.3 Fuzzy model

In Table 3, we present the estimation results for the new fuzzy approach, specified in the present paper. As in the results from JKL, AD and PD are the most important dimensions (the disutility of level 5). SC and UA are associated with largest imprecision of level-5 disutility (the difference between lower and upper disutility).

Due to the approach to estimation (defining constraints on and estimating the middles and lengths of $[l, h]$ intervals, rather than l and h), the parameters l

are non-increasing in several cases (UA5, PD5, AD5). The size of this effect is small, e.g. as compared to the estimation error (not presented here).

In this specification, the utility of the worst state, $u(55555)$, is an interval: $[-1.3; -0.57]$, much narrower than in JKL.

Most importantly, the measure of fit for the newly specified fuzzy approach clearly outperforms the two earlier specifications.

Table 3: Fuzzy model, dimensions as in Table 1, $\tau_1 = 0.406$, $\tau_2 = 1.229$, and $\tau_3 = 1.265$, $\omega = 0.293$. The measure of fit is 50392

Dimension	**Parameter**	**Level**			
		2	**3**	**4**	**5**
MO	l	0.022	0.033	0.234	0.300
	h	0.125	0.157	0.309	0.433
SC	l	0.000	0.000	0.167	0.247
	h	0.046	0.111	0.235	0.441
UA	l	0.080	0.036	0.265	0.181
	h	0.124	0.216	0.296	0.380
PD	l	0.060	0.064	0.473	0.459
	h	0.123	0.252	0.503	0.517
AD	l	0.044	0.183	0.404	0.382
	h	0.134	0.229	0.506	0.528

5 Discussion

We tested the quality of three approaches for modelling the preferences towards health states. We found that the FMN fuzzy-based approach specified in the present paper clearly outperforms the other two. On the other hand, the fuzzy-based approach suggested previously by JKL performed worst in terms of model fit. The relative differences between the measures of fit are quite large, as compared to the impact of adding/removing a constant term in the crisp specification.

These results suggest the following, in our opinion. Fuzzy modelling of preferences in the context of health not only has face validity (people find it difficult to introspectively determine their own preferences) but also performs better in terms of objective criteria. We attribute those difficulties to the fact that choosing between health states forces to consider conflicting objectives: respondents have to compare (a) different dimensions of health (e.g. juxtaposing mental and physical disabilities) and (b) different durations of remaining in a given health state. Still, it is very important to introduce fuzziness properly, so as to correctly model the imprecision and its impact on decisions.

JKL's idea to model the DCE data with fuzzy sets was correct, but the concrete specification can be improved. JKL based the probability of choosing a given alternative on the difference between the membership functions calculated along the Y axis. This approach had two important features: it directly corresponded to the constant proportional trade-off (CPTO) assumption (scaling the durations should not change the preferences between the alternatives) and allowed full certainty of choice when two alternatives differ substantially. Its poor performance may be due to the violation of CPTO in empirical data (Attema et al., 2010; Jakubczyk et al., 2017), and to the failure of the exponential Bradley-Terry function used in the crisp specification to follow the CPTO.

The new fuzzy-based specification resembles the crisp approach more directly, in that the utilities of the compared health states are subtracted (the middles of 1-cuts of utilities, to be precise) to derive the probability of one state being chosen. The fuzzy approach allows the imprecision in preferences to impact the behaviour of respondents in that the probability is shifted towards 50%. The improvement in model fit (along with estimated $\omega > 0$) suggests this mechanism may be in place.

We acknowledge that many improvements can still be made to all the specifications (especially to the crisp model, which contains fewer parameters), e.g. time can be handled non-linearly. Also, the approach to the estimation process (e.g. the monotonicity constraints) could be changed (e.g. to guarantee the monotonicity in rows in Table 3). Therefore, the result of the present comparison should not be treated as the ultimate test determining the correct approach, but rather to indicate the ideas to be pursued in subsequent research.

The individual results are quite consistent between the considered approaches: pain/discomfort and anxiety/depression are the most important dimensions (i.e. the worsening to level 5). In the new specification (preferred due to predictive validity over the JKL), the SC and UA dimensions are associated with especially large imprecision (particularly when measured relative to the disutility), i.e. the difference between h and l. This may result either from the fact that 'self care' and 'usual activities' are the most vague notions in the descriptive system, or from the fact that the respondents find it most difficult to assess the value of performing these activities. It might also be the case that the importance of these criteria varies with the duration (see Jakubczyk et al., 2017). Here, we would like to highlight that the obtained results show the ability of the fuzzy approach to capture this differing uncertainty of comparing conflicting criteria in a multiple-objective setting.

Our analysis provides insight into how time – yet another dimension in the multiple-criteria decision setting considered here – is perceived when comparing health states (in the crisp approach and the new fuzzy approach). The parameters τ measure the relative importance of a unit of time (relative to 'year' as a unit). All the shorter units (days, weeks, months) in both approaches were found to have larger weight than it would follow from their actual duration. For example, in the crisp approach, τ_1 (corresponding to 'day') amounts to 0.123 and in the fuzzy approach, to 0.406, while one day equals $1/365$ of a year. There are at least two possible interpretations. First, when the problem is presented in days, the decision maker changes his/her own attitude (the framing effect) and realizes that even individual days matter. The decision maker may pay attention to the relative, not only absolute, differences in duration between the alternatives (see Jakubczyk et al., 2017).

Second, when faced with a decision problem and overwhelmed with the amount of information about conflicting multiple objectives, the decision maker may focus on numbers (how many units of time will I live in this state) and not on units (what is the actual duration). In the extreme case, if the units are neglected altogether, we would expect all parameters τ to be equal to 1.

Based on our findings, in future research it would make sense to test other approaches to time handling. For example, the duration could enter the equations in non-linear form (see also Jakubczyk et al., 2017), to drop the CPTO assumption altogether. Also, other treatments of imprecision in eq. (8) could be considered: it is not obvious that the imprecisions should depend linearly on time in the same fashion that the utility does. No estimation errors were presented in the present study. A more systematic treatment of statistical significance of findings could help to direct further research. Finally, the comparison of models was based on the model fit. Perhaps the predictive validity (out-of-sample prediction success) could be a more reliable approach.

Acknowledgments

The research was financed by the funds obtained from National Science Centre, Poland, granted following the decision number DEC-2015/19/B/HS4/01729.

References

Attema A.E., Brouwer W.B. (2010), *On the (not so) Constant Proportional Trade-off in TTO*, Quality of Life Research, 19(4), 489-497.

Bansback N., Brazier J., Tsuchiya A., Anis A. (2012), *Using a Discrete Choice Experiment to Estimate Health State Utility Values*, Journal of Health Economics, 31, 306-318.

Beresniak A., Medina-Lara A., Auray J.P., De Wever A., Praet J.C., Tarricone R., Torbica A., Dupont D., Lamure M., Duru G. (2015), *Validation of the Underlying Assump-*

tions of the Quality-adjusted Life-years Outcome: Results from the ECHOUTCOME European Project, Pharmacoeconomics, 33(1), 61-69.

Bezanson J., Edelman A., Karpinski S., Shah V.B. (2017), *Julia: A Fresh Approach to Numerical Computing*, SIAM Review, 59, 65-98.

Bleichrodt H., Wakker P., Johannesson M. (1997), *Characterizing QALYs by Risk Neutrality*, Journal of Risk and Uncertainty, 15, 107-114.

Brooks R., De Charro F. (1996), *EuroQol: The Current State of Play*, Health Policy, 37, 53-72.

Du K.-L., Swamy M.N.S. (2016), *Search and Optimization by Metaheuristics*, Springer.

Gao F., Han L. (2012), *Implementing the Nelder-Mead Simplex Algorithm with Adaptive Parameters*, Computational Optimization and Applications, 51, 259-277.

Herdman M., Gudex C., Lloyd A., Janssen M., Kind P., Parkin D., Bonsel G., Badia X. (2011), *Development and Preliminary Testing of the New Five-level Version of EQ-5D (EQ-5D-5L)*, Quality of life research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 20(10), 1727-1736.

Jakubczyk M. (2009), *Impact of Complementarity and Heterogeneity on Health Related Utility of Life*, Central European Journal of Economic Modelling and Econometrics, 1, 139-156.

Jakubczyk M. (2015), *Using a Fuzzy Approach in Multi-criteria Decision Making with Multiple Alternatives in Health Care*, Multiple Criteria Decision Making, 10, 65-81.

Jakubczyk M., Craig B.M., Barra M., Groothuis-Oudshoorn C.G., Hartman J.D., Huynh E., Ramos-Goñi J.M., Stolk E.A., Rand-Hendriksen K. (2017), *Choice Defines Value: A Predictive Modeling Competition in Health Preference Research*, Value in Health, doi: 10.1016/j.jval.2017.09.016.

Jakubczyk M., Kamiński B. (2017), *Fuzzy Approach to Decision Analysis with Multiple Criteria and Uncertainty in Health Technology Assessment*, Annals of Operations Research, 251, 301-324.

Jakubczyk M., Kamiński B., Lewandowski M. (2017), *Eliciting Fuzzy Preferences Towards Health States with Discrete Choice Experiments* [in:] Studies in Systems, Decision and Control. Vol 125. Complex Systems: Solutions and Challenges in Economics, Management and Engineering, eds. C. Berger-Vachon, A.M.G. Lafuente, J. Kacprzyk, Y. Kondratenko, C.F. Merigó, Springer, 131-146.

Martí R. (2003), *Multi-Start Methods* [in:] *Handbook of Metaheuristics*, eds. F.W. Glover, G.A. Kochenberger, Springer, 355-368.

Menard S.W. (2002), *Applied Logistic Regression (2nd ed.)*, SAGE.

Miyamoto J.M., Wakker P.P., Bleichrodt H., Peters H.J.M. (1998), *The Zero-Condition: A Simplifying Assumption in QALY Measurement and Multiattribute Utility*, Management Science, 44(6), 839-849.

Pettitt D.A., Raza S., Naughton B., Roscoe A., Ramakrishnan A., Ali A., Davies B., Dopson S., Hollander G., Smith J.A., Brindley D.A. (2016), *The Limitations of QALY: A Literature Review*, Journal of Stem Cell Research & Therapy, 6(4), 334.

Weinstein M.C., Torrance G., McGuire A. (2009), *QALYs: The Basics*, Value in Health, 12 (S1), S5-S9.

White J.M., Mogensen P.K., Holy T., Riseth A.N., Lubin M., Stocker Ch., Ortner Ch., Johnson B., Noack A., Yu Y., Carlsson K., Lin D., Covert T.R., Rock R., Regier J., Kuhn B., Williams A., Ryan, Smith D., Anantharaman R., Gomez M., Revels J., Dunning I., MacMillen D., Rackauckas Ch., Legat B., Levitt A., Stukalov A., Petrov A., Mahajan A. (2017), *JuliaNLSolvers/Optim.jl*, doi: 10.5281/zenodo.1035790.

Zadeh L. (1965), *Fuzzy Sets*, Information and Control, 8(3), 338-353.

MULTIPLE CRITERIA DECISION MAKING
Vol. 12 2017

Paweł Tadeusz Kazibudzki*

NEW RESULTS ON THE QUALITY OF RECENTLY INTRODUCED INDEX FOR A CONSISTENCY CONTROL OF PAIRWISE JUDGMENTS

DOI: 10.22367/mcdm.2017.12.07

Abstract

A system exists which meets a prescription of the efficacious multiple criteria decision making support methodology. It is called the Analytic Hierarchy Process (AHP). The consistency control of human pairwise judgments about their preferences towards alternative choices appears to be the crucial issue in this concept. This research examines the efficiency of a recently proposed consistency index grounded on the redefined idea of triads inconsistency within Pairwise Comparison Matrices. The quality of the recently introduced proposal is studied and compared to other ideas with application of Monte Carlo simulations coded and run in Wolfram Mathematica 8.0.

Keywords: pairwise comparisons, consistency control, AHP, Monte Carlo simulations.

1 Introduction

It can be noticed that a world is a complex system of interacting elements. For instance, the contemporary economy depends mostly on energy. The availability of energy, on the other hand, depends on geography and politics but politics depends on military strength which depends on technology and access to energy. A technology depends on ideas, innovations and resources but ideas and innovations also depend on politics for their acceptance and support…, and so on. It is obvious that human minds have not yet evolved to the point where they can clearly perceive these ultimate relationships and solve crucial issues associated with them like for example nuclear energy, environmental regulations or global

* Jan Długosz University in Częstochowa, Department of Law, Administration and Management, Częstochowa, Poland, e-mail: p.kazibudzki@ajd.czest.pl.

crisis concerning third world poverty, population migration issues, society aging problems, etc. In order to deal with complex and fuzzy social, economic, and political issues, people must be supported and directed on their way to order priorities, to agree that one goal out-weighs another from a perspective of certain criterion, to make tradeoffs in order to be able to serve the greatest common interest.

Obviously, we cannot trust our intuition, although many of us commonly do it, devising solutions for complex problems which demand reliable answers. There are many examples showing that our intuition fails in such situations. Moreover, there are also many examples that our intuition fails anyway, even then when problems are relatively simple but their solution requires of involvement, not one, but two human's hemispheres.

Many examples exist indicating the fact that human's intuition misleads. There is a common riddle: a brick weighs a kilogram and a half of the brick. The question asks: what is a weight of the brick? For some reasons, a majority of people asked about it, although mathematical calculations are very trivial, provides the following incorrect answer: a brick weighs a kilogram and a half. It is presumably the principal reason why scientists continuously deal with explanation and modeling of decisional problems in the way they could be widely comprehended. That is why many supportive methodologies have been elaborated in order to make decision-making process easier, more credible and sometimes even possible.

An overwhelming scientific evidence indicates that the unaided human mind is simply not capable to analyze simultaneously many different competing factors and then synthesize them for the purpose of rational decision. Miller's well known experiment of 1956, titled, 'The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Information Processing' (Miller, 1956) made clear – almost a century ago – that the human mind is limited when considering short-term memory and discriminating skills of more than seven items. This indicates that when confronted with multiple variables, the choice made is less rational; and conversely, the less rational, the more alternatives available. This condition becomes more apparent when a choice is required from among several alternatives considered through a matrix of various criteria.

2 A methodology for decision making

An exceptionally popular tool designed especially to aid people in complex decision making, i.e. making a choice from various alternative based on a criteria matrix, is the 'Analytic Hierarchy Process' (AHP). The AHP seems to be the most widely used multicriteria decision making approach in the world today. The most recent list of application oriented papers one may want to find for instance in Grzybowski (2016). Actual applications in which the AHP results were accepted and used by the competent decision makers for instance can be found in: Saaty (2008), Ishizaka and Labib (2011), Ho (2008), Vaidya and Kumar (2006).

Currently the most popular method of assessing preferences regarding various decisional variations in an AHP is the 'Right Eigenvector Method' (REV). This approach takes advantage of information contained in the 'Pairwise Comparison Matrix' (PCM) which reflects the decision-maker's preferences expressed as linguistic variables – more or less fuzzy. Thus, it is possible to use words to compare qualitative factors and derive ratio scale priorities that can be combined with quantitative factors.

To make it possible a scale is utilized in order to evaluate the preferences for each pair of items. Supposedly, the most popular is Saaty's numerical scale which comprises the integers from one (equivalent to the verbal judgment: "equally preferred") to nine (equivalent to the verbal judgment: "extremely preferred") and their reciprocals. However, in conventional AHP applications we may want to utilize also other scales, i.e.: geometric scale and numerical scale. The first one usually consists of the numbers computed in accordance with the formula $2^{n/2}$ where n comprises the integers from minus eight to eight. The latter involves arbitrary integers from one to n and their reciprocals.

The first step in using AHP is to develop a hierarchy by breaking the problem down into its components. The basic AHP model includes goal (a statement of the overall objective), criteria (the factors one should consider in reaching the ultimate decision) and alternatives (the feasible alternatives that are available to reach the ultimate goal). Although the most common and basic AHP structure consists of a goal-criteria-alternatives sequence (Figure 1), AHP can easily support more complex hierarchies.

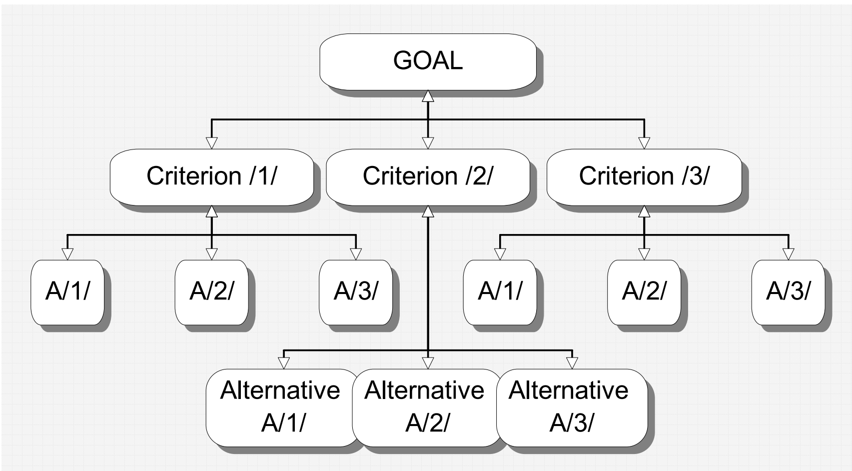


Figure 1. The most common exemplary hierarchy that consists of three levels: goal, three criteria, and three alternatives under each criterion

3 Introduction to the problem

One of the fundamental problems in AHP analysis is the priority weight assignment for the available decision alternatives. As it was stated earlier, the most popular method for estimating priority weights on the basis of the 'Pairwise Comparison Matrix' is the 'Right Eigenvector Method', proposed by Saaty and applied in the 'classic' AHP (Saaty, 1977). The conventional problem of AHP can be presented as:

$$\begin{bmatrix} x_1/x_1 & x_1/x_2 & x_1/x_3 & \dots & x_1/x_n \\ x_2/x_1 & x_2/x_2 & x_2/x_3 & \dots & x_2/x_n \\ x_3/x_1 & x_3/x_2 & x_3/x_3 & \dots & x_3/x_n \\ \vdots & \vdots & \vdots & & \vdots \\ x_n/x_1 & x_n/x_2 & x_n/x_3 & \dots & x_n/x_n \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} = \lambda_{\max} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} \tag{1}$$

and its outcome, i.e. the principal eigenvector $\boldsymbol{w} = [w_1,\dots, w_n]^T$ is provided by a solution of $\boldsymbol{X}\,\boldsymbol{w} = \lambda_{\max}\,\boldsymbol{w}$ where: $w_i > 0$, and $i = 1,\dots, n$.

Together with Saaty's method of priorities estimation, it was simultaneously proposed Saaty's 'Consistency Index'. What is important from the scientific point of view is that while the method contains several advantages, it also contains a series of very significant flaws which cannot be dismissed (Farkas, 2007).

It behooves mentioning those listed in literature on the subject, i.e. rank reversal, or the lack of any kind of quality criteria for the decision-maker to recognize why one decision vector weight is better than other evaluations. A significant drawback in the 'classic' approach of AHP is also the forced, reversed symmetry of the PCM which causes a loss of preference weight information contained in the elements of the 'ignored part' of a matrix (Grzybowski, 2012).

However, the most serious flaw of the AHP that was observed and stressed in current literature is the proposed, completely arbitrary method of recognizing (or not) the PCM as consistent enough for generating priority estimations (Grzybowski, 2012), and the very low correlation value between Saaty's sufficient consistency index values and the error value (absolute or relative) for the priority estimation weights (Grzybowski, 2016; Kazibudzki, 2016a). The examination of the latter issue is in order of this paper.

4 The problem description

It is obvious that even the best method of PVs estimation is useless until information about a scale of PCM inconsistency is provided. It is claimed and it is quite intuitive that serious errors in judgments about 'true' preferences of decision makers cause the data contained in PCM useless and result in poor estimates of decision makers' priorities (Saaty, 1980; Saaty, 2004; Saaty and Vargas, 1984). Therefore, we are presented with a number of papers dealing solely with the analysis of the inconsistency of the PCM. Undeniably, the consistency control and the evaluation of decision makers' inconsistency during the judgmental process is and should be a crucial part of every AHP study (Bulut et al., 2012; Aguaron et al., 2014; Altuzarra et al., 2010). The importance of the inconsistency control in the AHP practice was also emphasized in a number of application-oriented articles (Bulut et al., 2012; Pelaez and Lamata, 2003), group decision making oriented papers (Aguaron et al., 2014; Zhang et al., 2012), and research papers dedicated to elaboration of algorithms that lead to the consistency amelioration (Jarek, 2016; Xia et al., 2013; Benitez, 2012; Bozóki et al., 2011; Koczkodaj and Szarek, 2010).

In order to control the PCM consistency, different formulas (called indices) are proposed. These indices reflect in their way the degree of the PCM deviation from the one obtained in a perfect judgment case.

The first and the most popular inconsistency index (CI) was introduced by Saaty (1977) in his fundamental paper devoted to the AHP. His CI (denoted here as SI – formula 2) is closely related to the REV.

$$SI = \frac{\lambda_{\max} - n}{n-1} \tag{2}$$

The other popular CI is connected with a prioritization procedure (PP) that is known as the Row Geometric Mean method (GM) that was introduced by Crawford and Williams (1985) together with the Geometric Consistency Index (denoted here as GI – formula 3).

$$GI = \frac{2}{(n-1)(n-2)} \sum_{i<j} \log^2\left(\frac{x_{ij} w_j}{w_i}\right) \tag{3}$$

Another interesting concept of CI devised Koczkodaj (1993) who proposed his CI (denoted here as KI – formulae 4 and 5) that is based on the notions of a triad and its inconsistency.

$$KI(TI) = \max[TI(\alpha,\beta,\chi)] \tag{4}$$

$$TI(\alpha,\beta,\chi) = \min\left[\left|1 - \frac{\beta}{\alpha\chi}\right|, \left|1 - \frac{\alpha\chi}{\beta}\right|\right] \tag{5}$$

for: α, β, χ that are called a *triad*, where: $\alpha = a_{ik}$, $\chi = a_{kj}$, $\beta = a_{ij}$ for some different $i \le n, j \le n$, and $k \le n$, in a particular PCM denoted as: $\boldsymbol{A}(x) = [x_{ij}]_{nxn}$. It behooves mentioning that KI is not associated with any specific PP.

Apart from the indices SI, GI and KI, there exist and are promoted different other CI for PCMs, see for example: Kazibudzki (2016b), Dijkasra (2013), Grzybowski (2012). There are also some proposals for consistency control in the fuzzy pairwise comparison framework, such as the centric consistency index (which is based on GI) proposed by Bulut et al. (2012). However it seems undoubted, that these three above-mentioned indices (SI, GI and KI) are the most widely used ones in the pairwise comparisons methodology, see for instance Choo and Wedley (2004), Lin (2007), Grzybowski (2012), and Dong et al. (2008). All known from literature CI have one common characteristics, i.e. they are positive values and in the case of PCM perfect consistency they equal zero – what constitute a prerequisite of this theory. It is also believed that high CI values indicate poor consistency of decision makers' judgments what is supposed to indicate low quality of their preferences estimates. Obviously, such a belief is supported exclusively by some heuristic arguments which are mostly based on different intuitive psychological requirements, which according to the authors' opinions, should be reflected by CI properties.

It is important to underline that the most crucial and in the same time purely heuristic claim for common CI is the following assumptions: 'the more inconsistent judgments of decision makers are, the poorer are the estimates of priority weights'. Although it seems intuitive it turns out that it cannot be taken as granted (Grzybowski, 2016). Thus it is important to distinguish the following issues:

– the relation between the PCM consistency (reflected by CI) and the trustworthiness of decision makers judgments, and
– a dependence of the priority weights estimation errors from the level of PCM consistency designated by a given CI.

The pronounced majority of research devoted to inconsistency analysis, to our best knowledge except two papers, i.e. Grzybowski (2016) and Kazibudzki (2016a), as far combined the above mentioned issues and the existence of the distinguished earlier relations, i.e. among CI values, judgment consistency, and magnitudes of priority weights estimation errors, altogether treated as granted.

However, we should distinguish these two areas of study. The first, which can be perceived from the perspective of decision makers expertise (Brunelli and Fedrizzi, 2013) and the second, which defines the estimation quality of priority weights.

In this study we focus on the second problem, which constitute the primary research area of multicriteria decision making theory. We intend to study the relation between the values of CI and the magnitude of priority weights estimation errors. Thus, we are primarily interested in examination of the usefulness of the PCM as a source of information for estimation of priority weights. Hopefully, the results of our examination will allow decision makers to select such CI that is the most suitable from the perspective of their designated objectives.

5 The problem analysis

In order to examine a performance of selected CI from the assumed perspective, the following simulation scenario was considered. Its assumptions were introduced by Grzybowski (2016) then discussed and implemented in the paper of Kazibudzki (2016a). The simulation scenario comprises the following steps:

Step /1/ Randomly generate a priority vector $\boldsymbol{k} = [k_1,\ldots, k_n]^T$ of assigned size [n x 1] and related perfect PCM(k) = $\boldsymbol{K}(k)$.

Step /2/ Randomly choose an element k_{xy} for $x < y$ of $\boldsymbol{K}(k)$ and replace it with $k_{xy}e_B$ where e_B is relatively a significant error which is randomly drawn from the interval D_B with assigned probability distribution π.

Step /3/ For each other element k_{ij}, $i < j \leq n$ randomly choose a value e_{ij} for the small error in accordance with the given probability distribution π and replace the element k_{ij} with the element $k_{ij}\, e_{ij}$.

Step /4/ Round all values of $k_{ij}\, e_{ij}$ for i < j of $\boldsymbol{K}(k)$ to the closest value from a considered scale.

Step /5/ Replace all elements k_{ij} for $i > j$ of $\boldsymbol{K}(k)$ with $1/k_{ij}$.

Step /6/ After all replacements are done, calculate the value of the examined index as well as the estimates of the vector $\boldsymbol{k}$, denoted as $\boldsymbol{k}$*(EP), with application of assigned estimation procedure (EP). Then compute estimate errors AE($\boldsymbol{k}$*(EP), $\boldsymbol{k}$) and RE($\boldsymbol{k}$*(EP), $\boldsymbol{k}$) denoting the absolute and relative error respectively, where:

$$AE\left(k^*(EP),k\right)=\frac{1}{n}\sum_{i=1}^{n}\left|k_i-k_i^*(EP)\right|$$

$$RE\left(k^*(EP),k\right)=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{k_i-k_i^*(EP)}{k_i}\right|$$

Remember values computed in this step as one record.

Step /7/ Repeat Steps 2 to 6 N_M times.

Step /8/ Repeat Steps 2 to 7 N_R times.

Step /9/ Return *all* records organized as one database.

Source: Kazibudzki (2016a, p. 75).

The probability distribution π attributed in Step /3/ for e_{ij} is applied in equal proportions as: *gamma*, *log-normal*, *truncated normal*, and *uniform* distribution. The simulation scenario assumes that the factor e_{ij} is drawn from the interval $e \in [0,5;1,5]$ with the expected value of e_{ij} EV(e_{ij}) = 1. The 'big error' applied in Step /2/ has the uniform distribution on the interval $e_B \in [2;4]$.

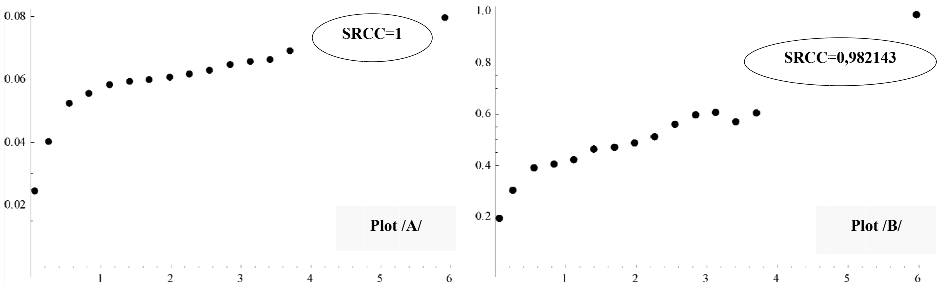
In this paper we examine simulation results for recently introduced by Kazibudzki (2016a) PCM consistency indicator – formula (6) with its yet not examined variation denoted by formula (7).

$$MLTI\left(LTI\right)=\frac{1}{N}\sum_{i=1}^{N}\left[LTI_i\left(\alpha,\beta,\chi\right)\right] \tag{6}$$

$$LTI\left(\alpha,\beta,\chi\right)=\ln^2\left(\alpha\chi/\beta\right) \tag{7}$$

where $LTI(\alpha,\beta,\chi)$ defines a seminal formula intended for indication of triad's consistency.

Due to necessity of diminishing the volume of the paper we present only results for n = 4 (where n denotes the number of alternatives in the model). For the same reason we implement only one estimation procedure, i.e. Logarithmic Least Squares Method (LLSM). Our simulation scenario assumes application of the rounding procedure which in this research operates according to Saaty's scale. Finally, our scenario takes into account the obligatory assumption in conventional AHP applications, i.e.: the PCM reciprocity condition. The results are presented in Tables 1-2 and Figure 2.



Note: SRCC stands for Spearman rank correlation coefficient.

Figure 2. Performance of the index MLTI(*LTI*). Plots of correlation between average values of MLTI(*LTI*) within analyzed interval and – the average AE (Plot A) and the average RE (Plot B)

In order to compare simulation results for MLTI(*LTI*) with performance of other consistency indices commonly used or proposed as good inconsistency indicators we present relations between fluctuations of selected consistency indices and selected characteristics of absolute or relative estimation errors. In order to save a space of the paper the results are only pictured on Figures 3-4.

Table 1: Performance of the index MLTI(*LTI*) in relation to AE(*LLSM*) distribution

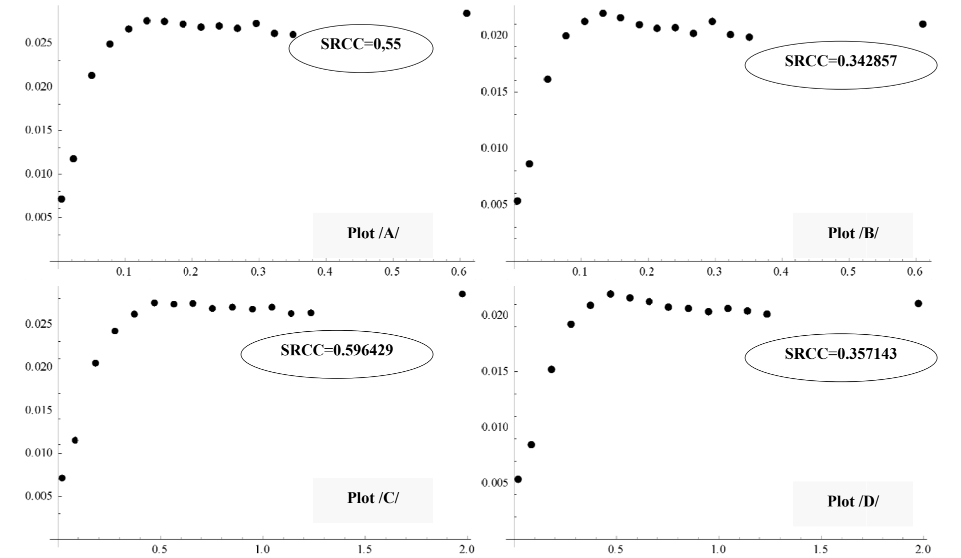
Average MLTI	***p*-quantiles of *AE*(LLSM)**			**Average *AE*(LLSM)**
	***p* = 0,1**	***p* = 0,5**	***p* = 0,9**	
0,05596	0,0071390	0,0176010	0,048430	0,0245701
0,25057	0,0114910	0,0309079	0,081851	0,0402469
0,54720	0,0204972	0,0467688	0,092151	0,0523990
0,83115	0,0241947	0,0490454	0,095879	0,0555646
1,12041	0,0262167	0,0531811	0,096393	0,0584429
1,40481	0,0275306	0,0552738	0,095133	0,0594058
1,68964	0,0273575	0,0553371	0,097423	0,0598936
1,97632	0,0274635	0,0555491	0,100479	0,0606786
2,26292	0,0268790	0,0559390	0,103806	0,0617115
2,55136	0,0270048	0,0565451	0,107156	0,0629664
2,84257	0,0267839	0,0570167	0,113131	0,0648082
3,12748	0,0270025	0,0579643	0,115005	0,0658326
3,41583	0,0262393	0,0594124	0,116590	0,0662670
3,70311	0,0263055	0,0614980	0,122258	0,0691538
5,92187	0,0285198	0,0721938	0,142200	0,0797634

Note: results based on 20 000 random reciprocal PCMs.

Table 2: Performance of the index MLTI(*LTI*) in relation to RE(*LLSM*) distribution

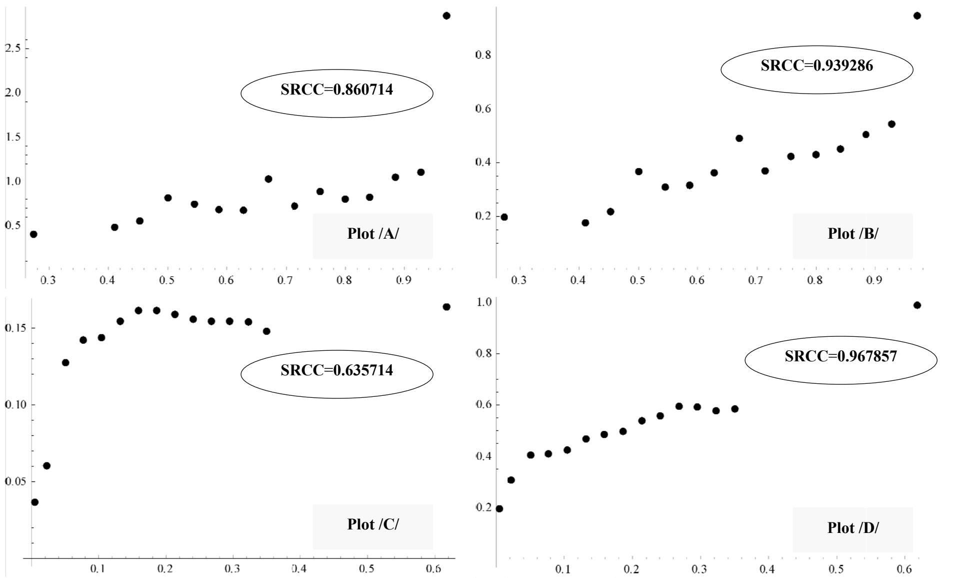
Average MLTI	***p*-quantiles of *RE*(LLSM)**			**Average *RE*(LLSM)**
	***p* = 0,1**	***p* = 0,5**	***p* = 0,9**	
0,05629	0,036686	0,083857	0,242085	0,195402
0,24889	0,058777	0,154707	0,463732	0,303042
0,54581	0,118697	0,233932	0,562547	0,391110
0,82979	0,141469	0,250504	0,571199	0,406388
1,11668	0,142768	0,274579	0,562745	0,423317
1,40159	0,151422	0,271206	0,551958	0,463763
1,68759	0,159168	0,266687	0,593866	0,471272
1,97225	0,162532	0,267185	0,624678	0,487771
2,26086	0,160420	0,274621	0,678569	0,511222
2,54802	0,157078	0,283623	0,716009	0,561110
2,83765	0,154172	0,289968	0,748193	0,598254
3,12254	0,154614	0,304405	0,764992	0,607791
3,41062	0,153110	0,308901	0,832981	0,570705
3,69985	0,150794	0,325909	0,835161	0,605517
5,96260	0,164129	0,393369	1,497960	0,987263

Note: results based on 10 000 random reciprocal PCMs.



Note: SRCC stands for Spearman rank correlation coefficient.

Figure 3. Performance of the indices SI and GI. Plots depict correlations between mean values of SI within analyzed interval and – AE quantiles of order 0,1 (Plot A) and AE quantiles of order 0,05 (Plot B), as well mean values of GI within analyzed interval and – AE quantiles of order 0,1 (Plot C) and AE quantiles of order 0,05 (Plot D)



Note: SRCC stands for Spearman rank correlation coefficient

Figure 4. Performance of the indices KI and SI. Plots depict correlations between mean values of KI within analyzed interval and – RE quantiles of order 0,95 (Plot A) and mean RE (Plot B), as well mean values of SI within analyzed interval and – RE quantiles of order 0,1 (Plot C) and mean RE (Plot D)

6 Discussion

It is believed that high CI values mean poor consistency of judgments what is supposed to entail low quality of decision makers' preferences estimates. This examination manifested that such a belief is supported exclusively by some heuristic arguments which according to some opinions, should be reflected by CI properties. The common assumption: 'the more inconsistent judgments of decision makers are, the poorer are the estimates of priority weights', cannot be taken as granted any more. Thus, we studied the relation between the values of selected CI and the magnitude of priority weights estimation errors.

We examined three commonly proposed inconsistency indicators for Pairwise Comparison Matrices, i.e. Saaty's consistency index (SI), geometric consistency index (GI), Koczkodaj's consistency index (KI), and the alternative proposition for consistency control, recently introduced by Kazibudzki (2016a), i.e. MLTI(*LTI*) index. We found out on the basis of analyzed cases that it is not true that a lower value of consistency index directly lead to a better estimation accuracy of decision makers' preferences. If that was true, we could observe a high and positive correlation between average values of selected consistency indices and relative or absolute estimation errors of simulated priority vectors. However, this research indicates that for GI, KI and SI, we can actually witness the situation when a decrease of consistency may lead to the improvement of a priority vector estimation quality, and inversely, when a growth of consistency may lead to the deterioration of a priority vector estimation quality (Figures 3-4). Our research indicates that in many analyzed cases we witness a non-monotonous relationship between values of a given consistency indicator and absolute or relative estimation errors of decision makers' preferences. However, it is not the case of proposed herein and examined new proposition for consistency control, i.e. MLTI(*LTI*) index – Figure 2, Tables 1 and 2. Its most serious advantages in comparison with other consistency indicators are: it is not connected with any prioritization procedure, it performs better than other analyzed consistency indicators and it can work also with AHP models that assume application of nonreciprocal PCM.

7 Conclusions

We have analyzed a performance of selected inconsistency indicators for simulated pairwise judgments from the perspective of their relations to absolute or relative estimation errors of decision makers' preferences.

We found out on the basis of analyzed cases that there exists a discrepancy between a common belief and a reality, i.e. it is not true that a lower values of consistency indicator directly lead to a better estimation accuracy of decision makers' preferences. It is a very important discovery because many authors still dedicate their research to the methods or procedures which strive to diminish some targeted consistency indicator.

Our research indicates that in many analyzed cases we witness a non-monotonous relationship between values of a given consistency indicator and absolute or relative estimation errors of decision makers' preferences. It means we should reform the concept of pairwise judgments consistency and search for such consistency indicators which reflect better the estimation quality of decision makers' priorities. It is so because the most commonly used consistency indicators may mislead about the estimation quality of decision makers' preferences.

The research indicates that in some cases we witness a situation when diminishing of a particular consistency indicator can lead to the deterioration of estimation quality. However it is certainly not a point of many researchers' effort. Thus, we should learn how to search and find new consistency indicators which possess features that are desired.

In this article we examined the consistency indicator that performs relatively well and it was recently introduced as a competitive solution for a consistency control of pairwise judgments.

References

Aguarón J., Escobar M.T., Moreno-Jiménez J.M. (2014), *The Precise Consistency Consensus Matrix in a Local AHP-group Decision Making Context*, Annals of Operational Research, DOI: 10.1007/s10479-014-1576-8.

Aguarón J., Moreno-Jiménez J.M. (2003), *The Geometric Consistency Index Approximate Thresholds*, European Journal of Operational Research, 147(1), 137-145.

Altuzarra A., Moreno-Jiménez J.M., Salvador M. (2010), *Consensus Building in AHP-Group Decision Making: A Bayesian Approach*, Operations Research, 58, 1755-1773.

Benítez J., Delgado-Galván X., Izquierdo J., Pérez-García R. (2012), *Improving Consistency in AHP Decision-making Processes*, Applied Mathematics and Computation, 219, 2432-2441.

Bozóki S., Fülöp J., Koczkodaj W.W. (2011), *An LP-based Inconsistency Monitoring of Pairwise Comparison Matrices*, Mathematical and Computer Modeling, 54(1-2), 789-793.

Brunelli M., Fedrizzi M. (2015), *Boundary Properties of the Inconsistency of Pairwise Comparisons in Group Decisions*, European Journal of Operational Research, 240(3), 765-773.

Bulut E., Duru O., Kececi T., Yoshida S. (2012), *Use of Consistency Index, Expert Prioritization and Direct Numerical Inputs for Generic Fuzzy-AHP Modeling: A Process Model for Shipping Asset Management*, Expert Systems with Applications, 39, 1911-1923.

Choo E.U., Wedley W.C. (2004), *A Common Framework for Deriving Preference Values from Pairwise Comparison Matrices*, Computers & Operations Research, 31, 893-908.

Crawford G., Williams C.A. (1985), *A Note on the Analysis of Subjective Judgment Matrices*, Journal of Mathematical Psychology, 29, 387-405.

Dijkstra T.K. (2013), *On the Extraction of Weights from Pairwise Comparison Matrices*, Central European Journal of Operations Research, 21(1), 103-123.

Dong Y., Xu Y., Li H., Dai M. (2008), *A Comparative Study of the Numerical Scales and the Prioritization Methods in AHP*, European Journal of Operational Research, 186, 229-242.

Farkas A. (2007), *The Analysis of the Principal Eigenvector of Pairwise Comparison Matrices*, Acta Polytech. Hung., 4(2), http://uni-obuda.hu/journal/ Farkas_10.pdf.

Grzybowski A.Z. (2016), *New Results on Inconsistency Indices and Their Relationship with the Quality of Priority Vector Estimation*, Expert Systems with Applications, 43, 197-212.

Grzybowski A.Z. (2012), *Note on a New Optimization Based Approach for Estimating Priority Weights and Related Consistency Index*, Expert Systems with Applications, 39, 11699-11708.

Ho W. (2008), *Integrated Analytic Hierarchy Process and Its Applications – A Literature Review*, European Journal of Operation Research, 186, 211-228.

Ishizaka A., Labib A. (2011), *Review of the Main Developments in the Analytic Hierarchy Process*, Expert Systems with Applications, 11(38), 14336-14345.

Jarek S. (2016), *Removing Inconsistency in Pairwise Comparison Matrix in the AHP*, Multiple Criteria Decision Making, 11, 63-76.

Kazibudzki P. (2016a), *Redefinition of Triad's Inconsistency and Its Impact on the Consistency Measurement of Pairwise Comparison Matrix*, Journal of Applied Mathematics and Computational Mechanics, 15(1), 71-78.

Kazibudzki P. (2016b), *An Examination of Performance Relations among Selected Consistency Measures for Simulated Pairwise Judgments*, Annals of Operations Research, 244(2), 525-544.

Koczkodaj W.W., Szarek S.J. (2010), *On Distance-based Inconsistency Reduction Algorithms for Pairwise Comparisons*, Logic Journal of the IGPL, 18(6), 859-869.

Koczkodaj W.W. (1993), *A New Definition of Consistency of Pairwise Comparisons*, Mathematical and Computer Modeling, 18(7), 79-84.

Lin CC. (2007), *A Revised Framework for Deriving Preference Values from Pairwise Comparison Matrices*, European Journal of Operational Research, 176, 1145-1150.

Miller G.A. (1956), *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Information Processing*, Psychological Review, 63, 81-97.

Peláez P., Lamata M. (2003), *A New Measure of Consistency for Positive Reciprocal Matrices*, Computers & Mathematics with Applications, 46, 1839-1845.

Saaty T.L. (2008), *Decision Making with the Analytic Hierarchy Process*, International Journal of Services Sciences, 1(1), 83-98.

Saaty T.L. (1977), *Scaling Method for Priorities in Hierarchical Structures*, Journal of Mathematical Psychology, 15(3), 234-281.

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

Saaty T.L. (2004), *Decision Making – The Analytic Hierarchy and Network Processes*, Journal of Systems Science and Systems Engineering, 13(1), 1-35.

Saaty T.L., Vargas L.G. (1984), *Comparison of Eigenvalue, Logarithmic Least Square and Least Square Methods in Estimating Ratio*, Mathematical Modeling, 5, 309-324.

Vaidya O.S., Kumar S. (2006), *Analytic Hierarchy Process: An Overview of Applications*, European Journal of Operation Research, 169, 1-29.

Xia M., Xu Z., Chen J. (2013), *Algorithms for Improving Consistency or Consensus of Reciprocal [0,1]-valued Preference Relations*, Fuzzy Sets and Systems, 216, 108-133.

Zhang G., Dong Y., Xu Y. (2012), *Linear Optimization Modeling of Consistency Issues in Group Decision Making Based on Fuzzy Preference Relations*, Expert Systems with Applications, 39, 2415-2420.

MULTIPLE CRITERIA DECISION MAKING

Vol. 12 2017

Dorota Kuchta[*]
Radosław Ryńca[**]
Yasmin Ziaeian[***]
Adam Grudziński[****]

MULTICRITERIA ASSESSMENT OF THE ACADEMIC RESEARCH ACTIVITY

DOI: 10.22367/mcdm.2017.12.08

Abstract

In this paper a network DEA approach to deal with efficiency assessment will be presented and applied to the assessment of performance of members of an academic faculty of Wroclaw University of Science and Technology. The purpose of this study is to propose a solution to the problem of multicriteria assessment of faculty members at universities, discussing at the same time its advantages and disadvantages in the context of the higher educational system in Poland.

Keywords: efficiency, performance, Data Envelopment Analysis, research assessment.

1 Introduction

Data Envelopment Analysis (DEA) is a "data oriented" non-parametric approach for evaluating the performance of Decision Making Units (DMUs) which convert multiple inputs into multiple outputs. DEA as a mathematical programming procedure computes a comparative ratio of outputs to inputs for each DMU,

[*] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wrocław, Poland, e-mail: dorota.kuchta@pwr.edu.pl.
[**] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wrocław, Poland, e-mail: radoslaw.rynca@pwr.edu.pl.
[***] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wrocław, Poland, e-mail: yasmin.ziaeian@pwr.edu.pl.
[****] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wrocław, Poland, e-mail: 015228420@student.pwr.edu.pl.

which is reported as the relative efficiency score. DEA has grown in a short period into a powerful, quantitative and analytical tool for measuring and evaluating efficiency and has been successfully applied in many contexts such as banks, schools, universities and other industries. The DMUs may be of various types: organizations, departments, projects, individuals… The reason why DEA is considered to be a good approach is that it requires minimal assumptions about how the inputs and outputs relate to each other. The result of its application is a relative efficiency of DMUs with respect to the given set of DMUs; therefore no global information is required, which is also an advantage (Charnes et al., 1978).

Originally, DEA dealt with one-stage production processes with no reference to the internal structure of the DMUs (with one set of inputs and one set of outputs only). But DEA can be also used to determine the efficiency of multi-stage processes, taking into account the internal structure of the DMUs, which indicates the flow of the intermediate inputs and outputs among the stages (Despotis et al., 2016).

The identification of the inputs and outputs for the assessment of DMUs is usually not easy, especially if we do not deal with a typical industry production process, where inputs are typical industrial resources and outputs, typical industrial products. The inputs should include all resources and other factors which impact the outputs. The outputs should reflect all useful outcomes which are important to the assessment of the DMUs or, in other words, the different criteria of the DMUs assessment.

Academic research is one of the most important activities in higher education and it consumes a large portion of state and business income (Athanassopoulos and Shale, 1997). It can be seen as an important element for determining the quality and performance of universities and other research institutions, as well as of employees of the individual university and research units. At the same time, there exists no generally accepted system of criteria of research activity assessment, either on the institutional or personal level (Woelert, 2015; Retzer and Jurasinski, 2009). That is why it is important to search for improvements and possible alternative solutions in this area.

The DEA model has already been used for the assessment of DMUs in the education and research context, for instance for educational and research organizations and units, research projects and individual researchers (Athanassopoulos and Shale, 1997; Meng et al., 2008; Despotis et al., 2015; Kuchta et al., 2016; Lee and Worthing, 2016). The inputs for the assessment of individual researchers include: the number of years of employment, salary, position etc., while for institutions: the number of academic staff and PhD students, wealth of the institution etc. Various outputs or evaluation criteria have been proposed in various national systems of research evaluation (Meng et al., 2008; Hicks, 2012; Lee, 2011; Ghinolfi et al., 2014; Lee and Worthing, 2016). Those criteria include: the number and quality of publications and citations, important research awards granted, invited talks at important conferences, patent commercialization, cooperation with established companies, significant consultant reports, setting up national standards, participation in editorial boards or in organizing committees of meetings and conferences, obtaining external research funding, achievements in educating master's degree and PhD degree holders etc.

Considering research activity assessment criteria without taking into account inputs, such as the number of years of experience for individual researchers or wealth of research institutions etc., may distort the results. Outputs do not come from nowhere, but are results of inputs and a "production" process. Thus the DEA method, which always tries to find inputs which influence the outputs, seems to be the right approach, not used at present at Polish universities. That is why the present paper proposes such an application.

The existing applications of the DEA model to research evaluation (Athanassopoulos and Shale, 1997; Meng et al., 2008; Despotis et al., 2015; Kuchta et al., 2016; Lee and Worthing, 2016) use either one-stage or two-stage models. One-stage models assume a single set of inputs and a single set of outputs and the solution of a single DEA model. A two-stage model may mean either:

- The formulation of two (or more) single-input and single-output DEA models (Meng et al., 2008): the first one for aggregated outputs, representing the main aspects of a research activity (e.g. publications – the aggregated output is the number of publications, grants – the aggregated output is the monetary value of grants completed, education of researchers – the aggregated output can be the total number of master's and PhD degree holders "produced" by a researcher or an institution) and the second one for a selected aspect of a research activity, taking into account not the aggregated output, but all individual outputs (e.g. for the aggregated aspect "publications" we can consider here the number of publications in various types of journals, in conference proceedings, monographs, citations etc.).
- The formulation of one DEA model reflecting the logical inner structure of inputs and outputs (some of which function as both outputs and inputs); two stages of the research process are considered (as in Figure 1, this is called "network DEA").
  - For example, in Lee and Worthing (2016), where the evaluation of research institutions is considered, the authors use two stages of the process. The first stage are "publications", with the number of academic staff and the institution's wealth as inputs and publication-related outputs, and the

second stage is “grant applications”, where the publication-related outputs from the first stage become the inputs, and the outputs are the number and value of grants obtained.

- Despotis et al. (2015) whose approach, explained in detail in Despotis et al. (2016), is adopted in this paper and presented in Figure 2, where individual researchers are evaluated, has as the first stage the “productivity” of a researcher (with inputs such as the time at the present position and salary and outputs related to publications since the appointment to the present position) and as the second one, the overall “impact/recognition” of the researcher. The latter represents “the impact that the research work of the individual has in academia and the recognition which the researcher has gained as a result of his work” (Despotis et al., 2015). In the second stage, inputs are equal to outputs from the first stage plus an external input (publications of the researcher before the appointment to the present position) and outputs are based on citations and other achievements, such as invited talks or important awards.

As mentioned above, we adopt the approach from Despotis et al. (2016), thus the network DEA, to an academic faculty at one of Polish universities. We discuss its practical advantages and disadvantages, as well as limitations imposed by both the Polish system of research evaluation and the information about researchers available at Polish universities, comparing the evaluation model and the results with those obtained for an academic faculty at a university in Greece (Despotis et al., 2015).

The paper proceeds as follows. First we present the network DEA method. Next, we apply the model to the selected academic faculty: the results of a pilot academic performance measurement are discussed. In the last section we draw conclusions.

2 The network DEA method

In this section, the network DEA approach is presented for two-stage processes. But before we proceed to the application of the network DEA, let us recall the basic DEA notion, that of efficiency (Charnes et al., 1978):

$$Efficiency = \frac{weighted\ sum\ of\ inputs}{weighted\ sum\ of\ outputs} \tag{1}$$

The idea of DEA is that each DMU, while being assessed, can choose its own weights for the weighted input and output sums, which are applied to all the DMUs, and allow the DMU being evaluated to obtain the maximal efficiency according to formula (1). The inputs are marked in Figure 1 with *X* and the outputs with *Y*. In the basic, one-stage DEA the internal structure of the DMUs, thus the two processes and *Z* inside the DMU in Figure 1 are not taken into account: the DMUs are treated as black boxes.

In Figure 1 we see, in fact, a two-stage process: the external inputs *X* enter the first stage of the process to produce the final output *Y* and the intermediate outputs-inputs Z are outputs for the first stage and inputs for the second stage.

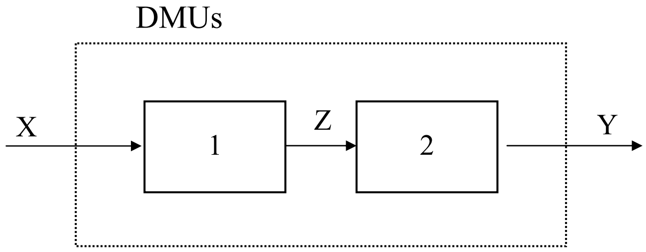


Figure 1. A two stage-process for network DEA

Source: Despotis et al. (2016).

In our model the two-stage process will be used (Figure 1).

Let us now specify the notation: assume *n* DMUs, indexed by $j = 1,\ldots, n$, each using *m* external inputs and producing *q* outputs, of the same nature for all the DMUs. The values of the inputs for the *j*-th DMU and the first stage are denoted as $X_j = \{x_{ij}, i = 1, \ldots, m\}$ and the values of the outputs for the same DMU and the same stage, as $Z_j = \{z_{pj}, p = 1, \ldots, q\}$. These outputs are used as inputs in the second stage, to produce *s* final outputs, whose values for the *j*-th DMU are denoted as $Y_j = \{y_{rj}, r = 1,\ldots, s\}$.

Let j_0 be a fixed index $j = 1,\ldots, n$. The efficiencies of the first and second stages of the j_0-th DMU are as follows (this is a direct consequence of formula (1)):

$$e_{j_0}^1 = \frac{\varphi^{j_0} Z_{j_0}}{\eta^{j_0} X_{j_0}}, e_{j_0}^2 = \frac{\omega^{j_0} Y_{j_0}}{\varphi^{j_0} Z_{j_0}} \tag{2}$$

In formulae (2), weights $\omega^{j_0} = \left(\omega_1{}^{j_0}, \ldots, \omega_s{}^{j_0}\right)$, $\eta^{j_0} = \left(\eta_1{}^{j_0}, \ldots, \eta_m{}^{j_0}\right)$, $\varphi^{j_0} = \left(\varphi_1{}^{j_0}, \ldots, \varphi_q{}^{j_0}\right)$ are used for inputs and outputs. Z_{j_0} occurs there in both functions, and in both cases has the same weights. The weights $\omega^{j_0} = = \left(\omega_1{}^{j_0}, \ldots, \omega_s{}^{j_0}\right)$, $\eta^{j_0} = (\eta_1{}^{j_0}, \ldots, \eta_m{}^{j_0})$, $\varphi^{j_0} = (\varphi_1{}^{j_0}, \ldots, \varphi_q{}^{j_0})$ are the values of the decision variables of mathematical programming problems (3), (4) and (5), where the efficiency of the j_0-th DMU form(s) the (maximized) objective functions.

We consider thus the following three ((3), (4) and (5)) mathematical programming problems with the j_0-th DMU in the main role:

$$\max e_{j_0}^1 = \frac{\varphi^{j_0} Z_{j_0}}{\eta^{j_0} X_{j_0}}$$
$$\text{s.t.}$$
$$\varphi^{j_0} Z_j - \eta^{j_0} X_j \leq 0, j = 1, \dots, n \tag{3}$$
$$\eta^{j_0} \geq 0, \varphi^{j_0} \geq 0$$

The optimal value of the objective function of problem (3) is the ideal efficiency of the first stage (Figure 1) for the j_0-th DMU and is denoted by $E_{j_0}^1$.

$$\max e_{j_0}^2 = \frac{\omega^{j_0} Y_{j_0}}{\varphi^{j_0} Z_{j_0}}$$
$$\text{s.t.}$$
$$\omega^{j_0} Y_j - \varphi^{j_0} Z_j \leq 0, j = 1, \dots, n \tag{4}$$
$$\omega^{j_0} \geq 0, \varphi^{j_0} \geq 0$$

Analogously to problem (3), the optimal value of the objective function of problem (4) is the ideal efficiency of the second stage (Figure 1) for the j_0-th DMU and is denoted by $E_{j_0}^2$.

The overall efficiency of the j_0-th DMU in the two-stage process from Figure 1 is defined (Despotis et al., 2016) using the solution of the following bicriteria problem:

$$\max e_{j_0}^1 = \frac{\varphi^{j_0} Z_{j_0}}{\eta^{j_0} X_{j_0}}$$
$$\max e_{j_0}^2 = \frac{\omega^{j_0} Y_{j_0}}{\varphi^{j_0} Z_{j_0}}$$
$$\text{s.t.}$$
$$\varphi^{j_0} Z_j - \eta^{j_0} X_j \leq 0, j = 1, \dots, n \tag{5}$$
$$\omega^{j_0} Y_j - \varphi^{j_0} Z_j \leq 0, j = 1, \dots, n$$
$$\eta^{j_0} \geq 0, \varphi^{j_0} \geq 0, \omega^{j_0} \geq 0$$

Of course, many approaches to the solution of the bicriteria problem (5) can be used. In Despotis et al. (2016) the distance of the solution of (5) from the ideal point $(E_{j_0}^1, E_{j_0}^2)$ is minimized, using a selected distance measure. Details can be found in Despotis et al. (2016). The final solution, thus the overall efficiency of the j_0-th DMU, is calculated by means of a single-criterion problem (Despotis et al., 2016). The solution is a vector $(e_{j_0}^1, e_{j_0}^2)$ representing the optimal solution of (5) following from the adopted assumptions, and the overall efficiency can be then expressed as the average or the product of the two values.

The model from Figure 1 can be also completed with an additional external input for Stage 2, denoted as L, which enters this stage together with the output of Stage 1 (as in Figure 2). Then the formula for $e_{j_0}^2$ changes to:

$$e_{j_0}^2 = \frac{\omega^{j_0} Y_{j_0}}{\varphi^{j_0} Z_{j_0} + \mu^{j_0} L_{j_0}} \tag{6}$$

and models (4), (5) and the other ones mentioned above, change accordingly Despotis et al. (2016).

3 Assessing the research activity of individual researchers by means of network DEA

The first attempt to use the model from Despotis et al. (2016) for the assessment of the research activity of individual researchers is Despotis et al. (2015), where a Greek university is analyzed. We will use an academic faculty at a Polish university as another case.

3.1 The case of a Greek university

In Despotis et al. (2015) the network DEA model from Despotis et al. (2016) was applied to the assessment of research activity of the researchers at an academic faculty of a university in Greece. As explained above, the network DEA model means that the process serving as the basis of the DMUs efficiency assessment is composed of two or more stages (Figure 1). In Despotis et al. (2015) the following model is used:

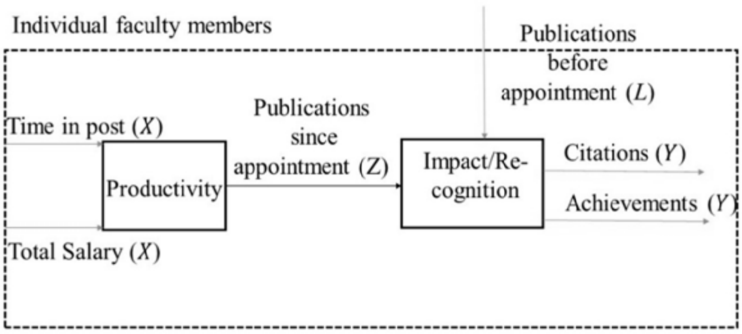


Figure 2. Academic research activity as a two-stage process

Source: Despotis et al. (2015).

The first stage represents the present (in the present position) productivity of researchers. The inputs in this stage are: time in the present position (such as full professor or associate professor) and the salary in the present position. Publications since appointment to the present position are the outputs of the first stage. The second stage represents the overall impact that the entire work of the researcher has had on science and the global recognition the researcher has gained in his/her entire academic career. In this stage the input consists of the output of the first stage and an external input, representing all the publications of the researcher before the appointment to the present position. Citations and important achievements (such as invited talks, important scholarly awards and positions etc.) are treated as final outputs. For the assessment of publications, the single-author equivalent (SAE) is used, which means that if, for example, a publication has three co-authors, each of them is assigned 1/3 of the publication.

The journals, and thus the publications in them, are either counted without any weighting (in one version of the model from Figure 2 in Despotis et al. (2015) or (in another version of the model from Figure 2 in the paper mentioned) classified in four quality classes (A+, A, B, C) according to the ERA2010 journal classification system[1] (www 1). A fifth class D is created for journals that are not indexed in ERA2010. The citations in both versions of the model were calculated as the number of units, without any weighting. The achievements (being editor-in-chief of a scholarly journal, associate editor or member of an editorial board, being invited as a keynote speaker to conferences, participating in organizing committees of conferences) were counted and weighted in a way which in Despotis et al. (2015) is not explained in detail.

The optimal values of the objective functions from the bicriteria model (5), representing the efficiencies of the researchers in Stages 1 and 2, were multiplied to form an overall efficiency of the researchers.

The most important conclusions from the case analyzed in Despotis et al. (2015) are the following:

- the difference between the case where the publications are counted without any reference to journal quality and the case where the journals, and thus publications, are classified according to the ERA system, is important: researchers with comparable efficiencies in the first case may have very different efficiencies in the second case;
- DEA delivers a better model for the evaluation of individual researchers than the conventional system of questionnaires, commonly used at universities at present;

[1] ERA: Excellence in Research in Australia.

- although DEA is a promising method for the evaluation of researchers, it is still unclear which model should be chosen; the following questions needs to be answered:
  - are the stages in Figure 2 defined correctly, i.e., so that they reflect the needs of the evaluation of researchers?
  - are the inputs and outputs defined correctly?
  - how to weight publications and citations, how to quantify achievements etc.?

The ERA system is just a proposal, and so are the quantifications of all the inputs and outputs used in Despotis et al. (2015).

3.2 The case of a selected academic faculty of a Polish university

The aim of the present paper is to test the approach from Despotis et al. (2015) in another context: at an academic faculty of a Polish university. The system presently used there for the assessment of researchers is based on questionnaires; information about the outputs in Figure 2 is requested, but without a distinction of stages, with the relevant information limited to the last 2-4 years. No information about inputs is used, at least formally. The journals are classified according to points assigned as explained below.

We adopted the model from Despotis et al. (2015) with the following problems and changes:

- We did not take the Australian ERA system for journal classification, but a system which is often used in Poland (www 2), based on a so-called "ministry list", elaborated and constantly updated by the Polish Ministry of Higher Education. Each paper in a journal included in this list is assigned a certain number of points (which may change from one year to another). The "ministry list" is composed of three parts:
  - A: journals with an impact factor and available in the JCR database (20 to 50 points);
  - B: journals that do not have an impact factor (0.25 to 10 points);
  - C: journals available in the European Reference Index for the Humanities (ERIH) database (12/16/20 points).
- We did not create any D class for journals not on the "ministry list". These journals were not taken into account; nor were book chapters or publications in conference proceedings not on the "ministry list". Of course, this decision may have influenced the results considerably. However, an initial study of the faculty in question has shown that it was publications in journals from the "ministry list" which really differentiated among the individual researchers.

- The information about achievements was not easily available. We had no access to the corresponding data. Therefore, we estimated the achievements on the basis of interviews and the incomplete knowledge we had of the persons in question, thus the results may be inaccurate in this regard. Additional input in the second stage was not considered.

As in Despotis et al. (2015), two versions of the model from Figure 2 are considered: one with all the publications from the "ministry list" treated equally (model M1) and another one with the number of points (according to the "ministry list" version from the year of the paper's publication) taken into account as weights (model M2).

The salary was not the actual one, to which we did not have access, but the average salary for the given position. Of course, both in Greece and in Poland, the researchers in question have also teaching duties, hence only a part of their salary is spent on research activities, but the exact percentage is impossible to determine with the Activity Based Cost approach, which is used only in selected English-speaking countries and in Scandinavia (Cropper and Cook, 2000).

Tables 1, 2 and 3 show the descriptive statistics of the data for the two models. Table 1 shows the years in the positions and the salaries; these data are used by both models.

Table 1: Descriptive statistics (part 1) of the data for Models M1 and M2

	Total Salary (in ten thousands)	**Time in Position (years)**
Min	2,78	0,49
Max	10,93	44,02
Average	5,44	21,91
St. Dev.	1,69	14,62

Table 2 shows the rest of the data for Model M1 and the output data shared by both models.

Table 2: Descriptive statistics (part 2) of the data for Model M1 (publications) and for both models (citations and achievements)

	Publications (SAE)	**Citations**	**Achievements**
Min	0	0	7
Max	59,67	409	34
Average	9,02	19,75	8
St. Dev.	12,29	77,73	7

Table 3 shows the data on publications for Model M2.

Table 3: Descriptive statistics (part 2) of the data on the publications for Model M2

	Publications (SAE)		
	A	B	C
Min	0,00	0,00	0,00
Max	4,33	40,08	5,25
Average	0,35	6,04	0,28
St. Dev.	1,02	7,64	1,00

It can be noticed that the number of publications in classes A and C is much smaller than that in class B. The results in almost all the categories are rather diversified, which is shown by the standard deviations and the differences between the maximal and minimal values.

Model (5) was solved. Figures 1 and 2 present the distribution of efficiencies of the researchers in the first stage of models M1 and M2.

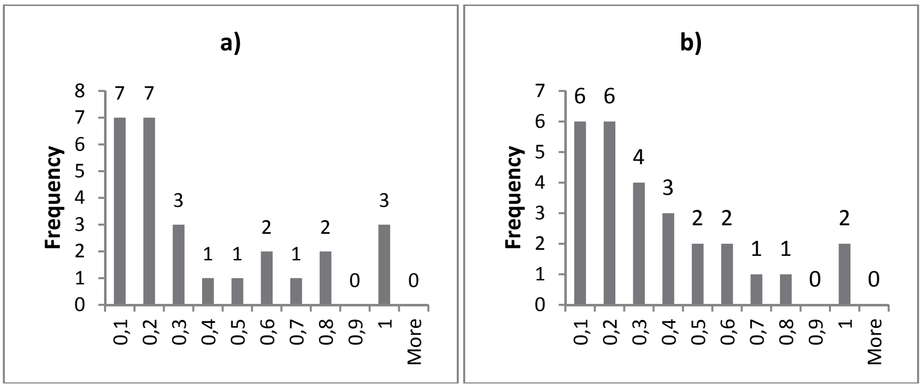


Figure 3. The distribution of productivities (efficiencies of the first stage, Figure 2) of researchers from the selected faculty of a Polish university for model M1 (a), all publications treated equally) and for model M2 (b), publications weighted according to journal quality)

A decrease in the productivity scores (Stage 1 in Figure 2) can be observed by comparing Figure 3a) (model M1, all publications treated equally) and 3b) (model M2, with the quality of publications taken into account). In the case of model M1 (Figure 3a)), the productivity of five researchers is over 0.7, in the other case (M2) only three researchers reach such a high productivity. This result can be visualized using the example presented in Table 4.

Table 4: Two selected researchers and their selected characteristics

	X	Y
Years in position	31.8	18.5
Total income in position (tens of thousands)	417.0	108.1
Publications after appointment (SAE total)	59.7	35.6
A (in SAE units)	2.1	0
B (in SAE units)	40.1	1
C (in SAE units)	5.3	0
Citations	409.0	1
Achievements	9	9
M1 − Productivity (Stage 1)	0.97	1
M2 − Productivity (Stage 1)	1	0.04

The two researchers can be regarded as similar with respect to the outputs, if the quality of journals is not taken into account. The mere fact of taking the quality of journals into account changes the assessment of the productivity of researcher B completely and the difference between the two researchers is visible.

Next we consider the second stage.

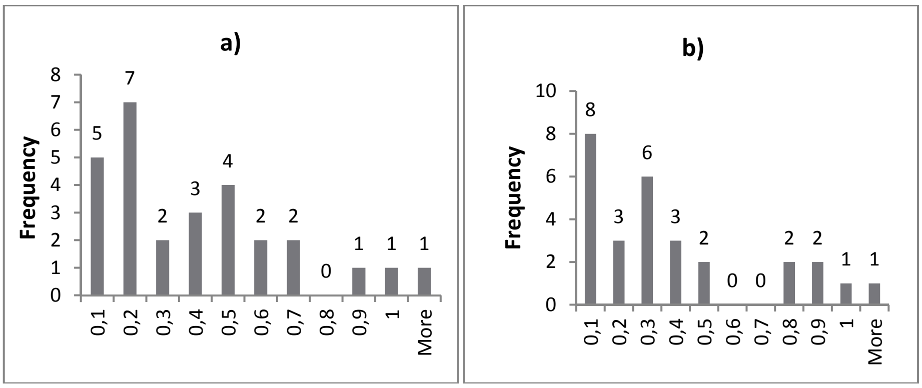


Figure 4. The distribution of impact/recognition (efficiencies of the second stage, Figure 2) of researchers from the selected faculty of a Polish university for model M1 (a), all publications treated equally) and for model M2 (b), publications weighted according to journal quality)

In the second stage of the model from Figure 2 the difference between not taking (Figure 4a)) and taking (Figure 4b)) journal quality into account has a lower influence on the results than in the first stage. It was probably because citations and achievements have a predominant impact on the results. It has to be kept in mind, however, that the citations were not weighed in either model and they come from journals of various quality and, as it was mentioned above, the information about achievements is here biased to a high degree. Hence, if the citations were weighted as the publications in Model M2 had been and if more exact information about achievement were available, the results in the two models in the second stage may have differed in a similar way they do in stage 1.

The overall efficiencies of the researchers are presented in Figure 5a) (model M1) and b) (model M2).

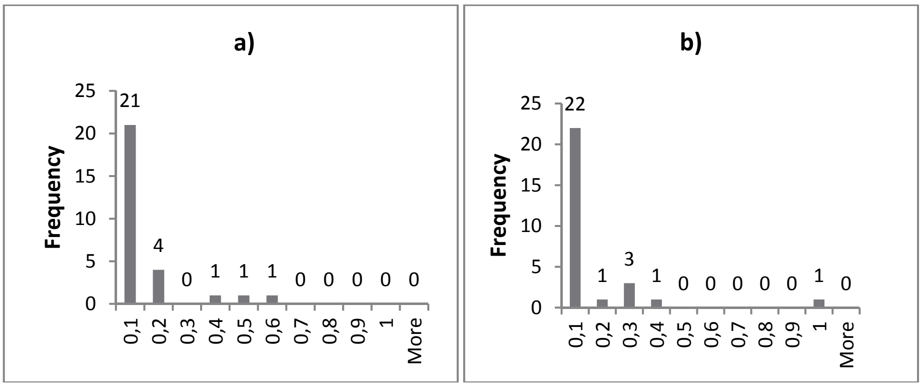


Figure 5. The overall efficiencies for both stages (Figure 2), calculated as a product, of researchers from the selected faculty at a Polish university for model M1 (a), all publications treated equally) and for model M2 (b), publication weighted according to journal quality)

As in Despotis (2015), the overall efficiency is calculated as the product of the values of both objective functions of the solution of problem (5). As a result, it is always relatively small and thus does not differentiate sufficiently among researchers. However, there is one researcher with the overall efficiency equal to 1 in model M2, hence he/she can be regarded as one who really distinguishes himself. Of course, in the analyzed case there is still the problem of insufficiently complete or exact data.

4 Conclusions

In our paper, a general network DEA approach to deal with efficiency assessment in a two-stage process has been presented, with multi-objective programming as the modeling framework. The proposed approach has been applied to an academic faculty of Wroclaw University of Science and Technology with 28 members. Research productivity is evaluated in the first stage, while the second stage represents the impact/recognition of individual researchers. Time in Position and Salary are the inputs of stage 1 and publications are its output. In stage 2, publications are the input, while achievements and citations are final outputs. Two versions of the model are considered: one with publications weighted equally and another one which takes into account their different quality.

It is important to underline the fact that the choice of criteria (outputs) we and other authors applying DEA use in assessing research activity is widely criticized (Retzer and Jurasinski, 2009; Vanclay, 2011; Woelert, 2015). First, it is emphasized that they are mainly quantitative and as such do not always reflect the actual value and output of a researcher or a research institution. Second, it has been shown that researchers, research institutions and academic journals have been adapting to the indicators so that higher values of the indicators do not correspond to a higher quality of research (Retzer and Jurasinski, 2009). There are attempts to include more qualitative and soft criteria in research evaluation (Retzer and Jurasinski, 2009), but this has not yet been incorporated into any DEA-based model. This step still needs to be done: DEA models will be useful only if the inputs and outputs used will correctly reflect the dependencies which exist in reality and the criteria which really do determine the quality of researchers.

Also, the network DEA model used in this paper has assumed a certain internal structure and stages in building up the output of the researchers. But the model is not ideal. First of all, the recognition of a researcher is a complex notion and it can be built up in another way than it was assumed in Despotis et al. (2015). In particular, non-quantitative data play here an important role, which is omitted in our model. The other problem is the multiplicative formula for the aggregated efficiency. In Despotis et al. (2015) and in our paper it gives, in most cases, very low values, hence the question arises whether it correctly differentiates among the researchers. Also, we can ask whether the input in the second stage (all the publications of the researcher being evaluated) is correct and what happens if he/she had very good publications in a remote past – is the model motivating in this case? Another question is how to separate the correct salary value corresponding to research activities, since the teaching load is very different in various institutions and countries.

The last important problem is the availability of data. In the case of the selected Polish university many important data are simply not available. Although publications are documented by library services and the corresponding information is available, citations are not classified according to any categories. This is not logical: if the weight of a publication depends on journal quality, so should the weight of a citation. And information about salary and achievements are sensitive data, which is protected and not available for scientific analysis of any kind. A coding method should be designed which would allow to store these data and make them available in a legally acceptable form. The Greek university is much smaller and made its data available in a legal form.

Also, there is the issue of the possible inclusion of publications in journals which are not on any “list”, of chapters in monographs or publications in conference proceedings. There are important contributions among them, which are valued very low − the second model in Despotis et al. (2015) or not at all (our model). This question still awaits an answer.

Of course, much more cases should be considered before a final recommendation can be made for universities on how to evaluate researchers or for governments on how to evaluate research institutions. However, it seems that DEA is, on the whole, a correct approach and it should replace the present system of researchers evaluation (and research institutions evaluations) which does not take inputs into account. Output cannot be analyzed disregarding the input, which is clearly shown by the example in Table 4 (the two researchers there differ strongly in experience and salary, which has to be taken into account in their analysis). But the question should be answered how to choose and measure inputs and outputs in individual stages and globally.

References

Athanassopoulos A.D., Shale E. (1997), *Assessing the Comparative Efficiency of Higher Education Institutions in the UK by the Means of Data Envelopment Analysis*, Educ. Econ, 5, 117-134.

Charnes A., Cooper W.W., Rhodes E. (1978), *Measuring the Efficiency of Decision Making Units*, European Journal of Operational Research, 2, 429-444.

Cropper P., Cook R. (2000), *Activity-Based Costing in Universities − Five Years On*, Public Money & Management, 20, 61-68.

Despotis D.K., Koronakos G., Sotiros D. (2015), *A Multi-objective Programming Approach to Network DEA with an Application to the Assessment of the Academic Research Activity*, Procedia Computer Science, 55, 370-379.

Despotis D.K., Sotiros D., Koronakos G. (2016), *A Network DEA Approach for Series Multi-stage Processes*, Omega, 61, 35-48.

Ghinolfi D., El Baz H.G., Borgonovin E., Radwan A., Laurence O., Sayed H.A., De Simone P., Abdelwadoud M., Stefani A., Botros S.S., Filipponi F. (2014), *A Model for Southern Mediterranean Research Institute Self-assessment: A SWOT Analysis-based Approach to Promote Capacity Building at Theodor Bilharz Research Institute in Cairo (Egypt)*, Arab Journal of Gastroenterology, 15(3-4), 92-97.

Hicks D. (2012), *Performance-based University Research Funding Systems*, Research Policy, 41(2), 251-261.

Kuchta D., Skorupka D., Duchaczek D., Kowacka M. (2016), *Modified, Stakeholders Perspective Based DEA Approach in IT and R&D Project Ranking*, ICEIS 2016: Proceedings of the 18th International Conference on Enterprise Information Systems, Rome, Italy, 25-28 April 2016, S. Hammoudi (ed.), 2, 158-165.

Lee B.L. (2011), *Efficiency of Research Performance of Australian Universities: A Reappraisal Using a Bootstrap Truncated Regression Approach*, Economic Analysis and Policy, 41(3), 195-203.

Lee B.L., Worthington A.C. (2016), *A Network DEA Quantity and Quality-orientated Production Model: An Application to Australian University Research Services*, Omega, 60, 26-33.

Meng W., Zhang D., Qi L., Liu W. (2008), *Two-level DEA Approaches in Research Evaluation*, Omega, 36, 6, 950-957.

Retzer V., Jurasinski G. (2009), *Towards Objectivity in Research Evaluation Using Bibliometric Indicators – A Protocol for Incorporating Complexity*, Basic and Applied Ecology, 10(5), 393-400.

Vanclay J.K. (2011), *An Evaluation of the Australian Research Council's Journal Ranking*, Journal of Informetrics, 5(2), 265-274.

Woelert P. (2015), *The 'Logic of Escalation' in Performance Measurement: An Analysis of the Dynamics of a Research Evaluation System*, Policy and Society, 34(1), 75-85.

(www 1) http://guides.library.unisa.edu.au/c.php?g=169983&p=1119054 (access: 13.05.2017).

(www 2) http://www.bg.pw.edu.pl/index.php/en/indicators-for-journals-evaluation#the-polish-ministry-of-science-and-higher-education-scores (access: 17.05.2017).

M U L T I P L E   C R I T E R I A   D E C I S I O N   M A K I N G
Vol. 12 2017

Aleksandra Łuczak*

ASSESSING THE STRATEGIC FACTORS AND CHOOSING THE DEVELOPMENT SCENARIOS FOR LOCAL ADMINISTRATIVE UNITS USING AHP

DOI: 10.22367/mcdm.2017.12.09

Abstract

The main aim of this study is to assess the strategic factors (objectives, tasks and development scenarios) and to select the best scenario for local administrative units. Two approaches are used to solve this problem: classical and fuzzy analytic hierarchy process based on experts' opinions. The research was based on data from surveys with the councillors of the urban and rural municipality of Międzychód and the rural municipality of Chrzypsko Wielkie. The importance of strategic factors for both municipalities was assessed, and the best development scenario was selected. As shown by the research, the most important scenario for the municipality of Chrzypsko Wielkie involves the development through support for entrepreneurship and agri-food processing, while that for the municipality of Międzychód involves the development by supporting housing, services and tourism.

Keywords: strategic factors, choice of scenario, AHP, FAHP.

1 Introduction

Local development planning involves solving many complex decision-making issues based on multiple criteria. One of them is to establish a development strategy which includes assessing the strategic factors (objectives, tasks and development scenarios) and choosing the best scenario. As the local development planning implies complex problems which require multidisciplinary know-how, it is helpful to rely on experts' opinions and on multiple-criteria decision-making

* Poznań University of Life Sciences, Faculty of Economics and Social Sciences, Department of Finance and Accounting, Poznań, Poland, e-mail: aleksandra.luczak@up.poznan.pl.

methods. Many various multiple-criteria decision-making methods exist, including the following groups: additive methods, analytic hierarchy and related methods, verbal methods, ELECTRE, PROMETHEE, reference point methods and interactive methods (see Trzaskalik, 2014a; 2014b). Each of these support methods features specific advantages, disadvantages and limitations. Therefore, it is important to select the right method for a specific decision-making problem. As regards local development planning, the problem is to assess the strategic factors and select the best development scenario among a finite number of options. This process consists in establishing the hierarchy of strategic factors, and ultimately results in selecting the best scenario. Thus, the decision makers are able to identify the most important strategic factors which, based on available information and preferences, may be regarded as the most appropriate ones and be recommended for use in local development planning. When solving such a problem, it is useful to employ the analytic hierarchy methods, including the classic (Saaty, 1980) and the fuzzy (Chang, 1996) analytic hierarchy process.

The aim of this study is to assess the importance of objectives, tasks and development scenarios, and to select the best scenario for local administrative units. To solve this problem, the assessment of strategic factors was underpinned by two approaches: the classic and the fuzzy analytic hierarchy process, which were based on opinions from municipal councillors. The surveys were conducted among the councillors of the Międzychód urban and rural municipality and of the Chrzypsko Wielkie rural municipality (Śmigielska, 2013).

2 Research methodology

The following steps are identified in the process of assessing the strategic factors and selecting the development scenario based on the classic (AHP) and fuzzy analytic hierarchy process (FAHP):

Step 1. Building the hierarchic diagram of strategic factors impacting the development of a local administrative unit.

Step 2. Pairwise comparisons of relevance between strategic factors.

Step 3. Validating the comparisons made by experts.

Step 4. Calculating the local and global priorities of strategic factors. Selecting the best development scenario.

The first step consists of building the hierarchic decision-making diagram which includes the key strategic factors impacting the development of a local administrative unit (Figure 1).

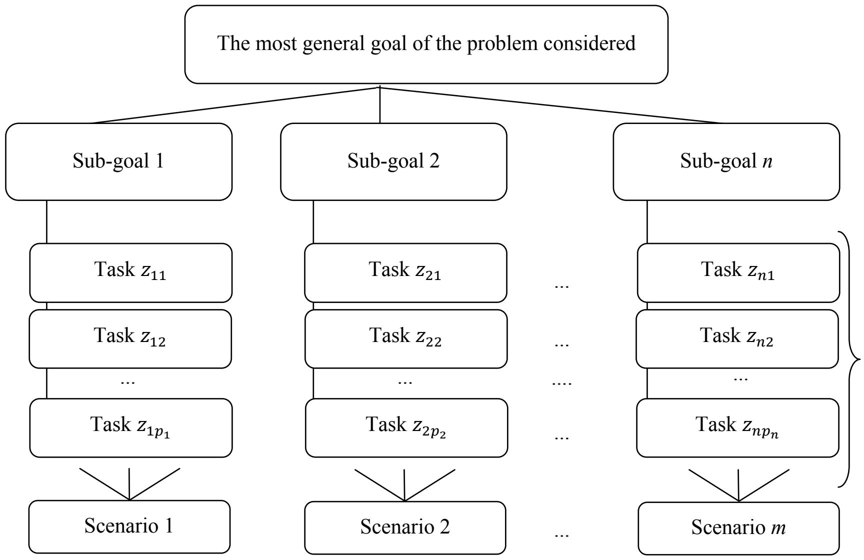


Figure 1. Hierarchic diagram of strategic factors for a local administrative unit

Source: Author's own study based on Saaty (1980).

Table 1: Nine-grade scale of preferences between two strategic factors

Intensity of importance	**Verbal description (assessment) of preferences**	**Intensity of importance, numerical assessment**	
		AHP	**FAHP**
Equal importance	Both factors contribute equally to achieve the objective	1	$\tilde{1}$ = (1, 1, 1)
Weak	Slight preference of one strategic factor over another (the first strategic factor is slightly more important than the other)	3	$\tilde{3}$ = (1, 3, 5)
Strong	Strong preference of one strategic factor over another (the first strategic factor is significantly more important than the other)	5	$\tilde{5}$ = (3, 5, 7)
Very strong	Very strong preference of one strategic factor over another (the first strategic factor is definitely more important than the other)	7	$\tilde{7}$ = (5, 7, 9)
Absolute	Absolute preference of one strategic factor over another (the first strategic factor is absolutely more important than the other)	9	$\tilde{9}$ = (7, 9, 9)
Compromise comparisons between the above values	If the decision maker is unable to choose between two neighboring ratings, intermediate values are used	2, 4, 6 and 8	$\tilde{2}$ = (1, 2,4); $\tilde{4}$ = (2, 4, 6); $\tilde{6}$ = (4, 6, 8); $\tilde{8}$ = (6, 8, 9)
Transitiveness of grades	If strategic factor i has one of the above grades assigned to it when compared with strategic factor j, then j has the reciprocal value when compared with i	reciprocals of above	reciprocals of above

Source: Saaty (1980); Wang et al. (2009).

The main strategic goal is placed at the top of the hierarchy (level I), and is broken up into sub-goals (level II). Each sub-goal includes separate packages of strategic tasks (level III) which affect the achievement of sub-goals. The tasks may also be decomposed into sub-tasks. The lowest level consists of the alternative decisions (development scenarios). The number of hierarchy levels depends on the intended level of detail (or generalization) to be maintained in the study, but it should not exceed seven[1]. Also, there should be no more than nine[2] sub-goals, tasks within a sub-goal or development scenarios.

In the second step, strategic factors are assessed by experts through pairwise comparisons of importance at each level of hierarchy. At level II, sub-goals are compared by their contribution (importance) to the main goal. At level III, the importance of strategic tasks which contribute to the sub-goals is compared. The lowest level consists in the pairwise comparison of development scenarios based on their impact on the implementation of specific strategic tasks. The nine-grade scale of preferences between two strategic factors is used in the comparisons (Table 1). Verbal descriptions of preferences are converted into real numbers in AHP, or into triangular fuzzy numbers in FAHP. The results are written down in the pairwise comparison matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1(\bullet)} \\ \dfrac{1}{x_{12}} & 1 & \cdots & x_{2(\bullet)} \\ \vdots & \vdots & \cdots & \vdots \\ \dfrac{1}{x_{1(\bullet)}} & \dfrac{1}{x_{2(\bullet)}} & \cdots & 1 \end{bmatrix},$$

where (•) stands for: n (the number of sub-goals), p_k (the number of tasks for k-th goal, $k = 1, 2, \ldots, n$), m (the number of scenarios); x_{ij} is the intensity of importance of strategic factor i over strategic factor j in AHP ($i, j = 1, \ldots, (\bullet)$); these are average values[3] of pairwise comparison assessments made by experts. Moreover, we have: $x_{ij} = 1/x_{ji}$ (transitiveness of grades) and $x_{ii} = 1$ (equivalence of grades). In the case of FAHP, x_{ij} ($i, j = 1, \ldots, (\bullet)$) are replaced by triangular fuzzy numbers $\tilde{x}_{ij} = (l_{ij}, m_{ij}, u_{ij})$, where $\tilde{x}_{ij} = 1/\tilde{x}_{ji} = (1/u_{ji}, 1/m_{ji}, 1/l_{ji})$ and $\tilde{x}_{ii} = (1,1,1)$.

In the third step, the pairwise comparisons are validated (see also Alonso and Lamata, 2006). For that purpose, the inconsistency ratio CR (Saaty, 1980) can be used:

[1] This is because of the limitations of short-term memory (Miller, 1956).

[2] This is because of the limited ability to memorize information: a human can compare a maximum of five to nine items without making any significant errors, depending on the type of information compared (Miller, 1956).

[3] The geometric mean or median can be used to average the experts' assessments.

$$CR = CI / RI \cdot 100\%,$$

where $CI = \frac{\lambda_{\max} - n}{n-1}$ is the inconsistency index; $\lambda_{\max}$ − the maximum or main eigenvalue of the comparison matrix **A**; n − the number of rows (columns) in matrix **A**; RI − the average random inconsistency index calculated based on a randomly generated $n \times n$ matrix[4] (Table 2).

Table 2: Average random inconsistency index

Matrix rank	*n*	1	2	3	4	5	6	7	8	9
Inconsistency index	*RI*	0.0000	0.0000	0.5245	0.8815	1.1086	1.2479	1.3417	1.4056	1.4499

Source: Alonso and Lamata (2006).

The inconsistency ratio CR should not exceed 10%. Otherwise, the information on preferences obtained from the experts needs to be verified as it suggests an excessive incoherence of pairwise comparisons between strategic factors. In that case, it is recommended to repeat the pairwise comparisons (Saaty, 1980).

As regards FAHP, if the pairwise comparisons are consistent in matrix **A** composed of elements m_{ij}, then the pairwise comparisons are also consistent in the fuzzy matrix $\widetilde{\mathbf{A}}$ (Csutora and Buckley, 2001).

The purpose of AHP and FAHP is to determine the scale vector based on pairwise comparisons, that is, to calculate local and global priorities (step 4). The method for calculating local priorities is illustrated by the example of tasks.

In the classic analytic hierarchy process (AHP), Saaty (1980) proposed the eigenvector method as a way to determine the values of local priorities (which can be approximated using various methods). The most widely adopted methods for determining the values of local priorities include the geometric mean. However, caution is recommended when using it, as the results may be inaccurate if pairwise comparison matrices are larger than 3 × 3. Another example is the normalized arithmetic mean of rows:

Step 1a. The pairwise comparison values in matrix **A** are normalized (normalization of columns):

$$s_{ij} = x_{ij} / \sum_{i=1}^{p_k} x_{ij} ,$$

where $i, j = 1, 2, \ldots, p_k$ (the number of tasks for k-th goal, $k = 1, 2, \ldots n$).

---

4 For other suggested RI values, see e.g. Saaty (1980), Aguaron and Moreno-Jiménez (2003), Alonso and Lamata (2006), Franek and Kresta (2014). RI values differ depending on the number of simulations. In studies conducted by Alonso and Lamata [2006], RI values are based on 100,000 matrices for each dimension. They also demonstrated that there were no significant differences between RI values in the case of 100,000 and 500,000 simulations.

Step 2a. Calculating vector w_i by averaging the values in the rows of **A**:

$$w_i = \frac{\sum_{j=1}^{p_k} s_{ij}}{p_k}.$$

The calculated values w_i are the local priorities.

In the fuzzy analytic hierarchy process (FAHP) proposed by Chang (1996), the values of local priorities are calculated as follows:

Step 1b. Calculating the fuzzy sum for each row of the fuzzy pairwise comparison matrix $\widetilde{\mathbf{A}}$ and normalizing them using operations on fuzzy numbers:

$$\widetilde{Q}_i = (l_i, m_i, u_i) = \frac{\sum_{j=1}^{p_k} (l_{ij}, m_{ij}, u_{ij})}{\sum_{i=1}^{p_k} \sum_{j=1}^{p_k} (l_{ij}, m_{ij}, u_{ij})}, \; i = 1, 2, \ldots, p_k;\; k = 1, 2, \ldots, n.$$

Step 2b. Calculating the degree of possibility that $\widetilde{Q}_i \geq \widetilde{Q}_g$ $(i, g = 1, 2, \ldots, p_k, i \neq g)$, as per the following formula:

$$V(\widetilde{Q}_i \geq \widetilde{Q}_g) = \text{hgt}(\widetilde{Q}_i \cap \widetilde{Q}_g) = \begin{cases} 1, & for \;\; m_i \geq m_g \\ 0, & for \;\; l_g \geq u_i \\ \dfrac{l_g - u_i}{(m_i - u_i) - (m_g - l_g)} & \text{otherwise,} \end{cases}$$

and selecting the minimum of the above values: $w_i^S = \min V(\widetilde{Q}_i \geq \widetilde{Q}_g)$. Upon normalization, the values w_i^S become the local priorities of tasks for p_k $(k = 1, 2, \ldots, n)$ as a part of sub-goal k:

$$w_i = w_i^S / \sum_{i=1}^{p_k} w_i^S.$$

Local priorities w_i for sub-goals[5] and development scenarios[6] are calculated similarly, as in steps 1a and 2a (for AHP) or as in steps 1b and 2b (for FAHP). Local priorities of levels II and III represent the contribution of the given strategic factor (sub-goal and task) to the goal of the next higher level. Local priorities are the basis for calculating global priorities w_i^g which represent the contribution of each strategic factor (on individual levels) to the main goal. The global priority is calculated by multiplying the local priority value of the strategic fac-

[5] Local and global priorities for each sub-goal are identical.
[6] Local priorities of development scenarios are calculated for each task.

tor at each hierarchy level by the global priority value of the related strategic factor at the next higher level.

At the lowest (scenario) hierarchy level, local priorities are multiplied by the corresponding global priorities of tasks. The results, referred to as "partial global priorities", illustrate the contribution of a scenario to the main goal through the implementation of the given task. The sum of all partial global priorities of a scenario is the global priority of that scenario. The values of global priorities for individual scenarios form the basis for selecting the best development scenario. The development scenario with the highest global priority is regarded as the most appropriate one.

3 Results of empirical studies

This study is an attempt to assess the strategic factors (goals, tasks and development scenarios) for the Międzychód urban and rural municipality and of the Chrzypsko Wielkie rural municipality (located in Międzychód county) using the classic and fuzzy analytic hierarchy processes. The first step was the establishment of the hierarchy of strategic factors[7] affecting the development of municipalities (Figure 2).

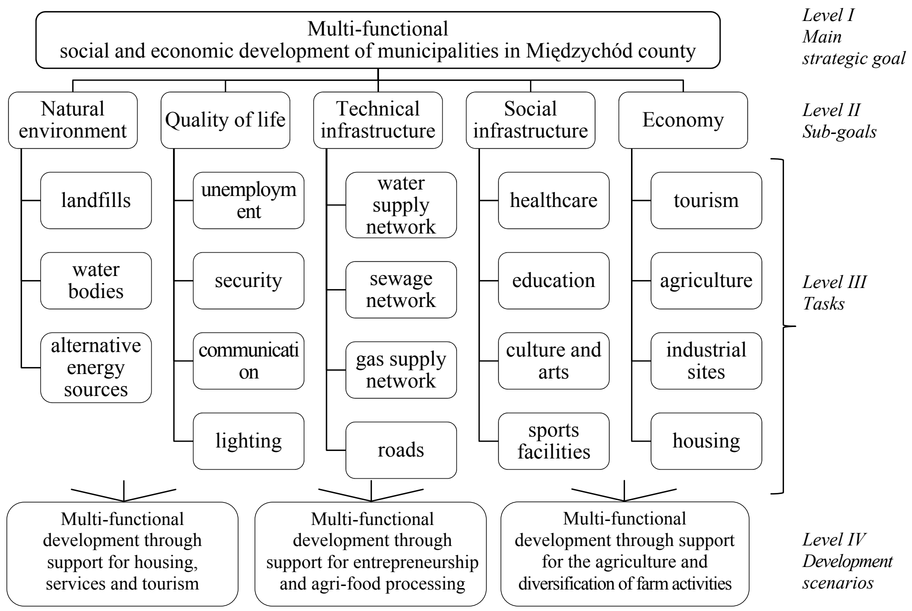


Figure 2. Hierarchy of strategic factors affecting the development of municipalities in Międzychód county

Source: Author's own study based on planning and strategic documents of municipalities in Międzychód county.

[7] The hierarchy includes the most important strategic factors for all municipalities in Międzychód county. One general hierarchy is necessary for comparing strategic factors in different types of municipalities.

The main strategic goal for the municipalities was assumed to be the multifunctional socio-economic development. To achieve the main goal, it is necessary to implement five sub-goals related to: the natural environment, the population's quality of life, the technical infrastructure, the social infrastructure and the local economy. Each sub-goal includes identified packages of strategic tasks which are essential for the attainment of sub-goals.

The following task packages were identified within the sub-goals:

Sub-goal 1: environmental protection (natural environment[8])

Tasks:
- solving the waste disposal issues (landfills),
- use of water bodies in the fisheries industry (water bodies),
- use of alternative energy sources (alternative energy sources).

Sub-goal 2: improving the population's quality of life (quality of life)

Tasks:
- reducing unemployment (unemployment),
- improving the population's sense of security (security),
- extending and upgrading the communication system (communication),
- improving the street lighting (lighting).

Sub-goal 3: expansion and modernization of the technical infrastructure (technical infrastructure)

Tasks:
- expansion and modernization of the water supply network (water supply network),
- expansion and modernization of the sewage network (sewage network),
- expansion and modernization of the gas supply network (gas supply network),
- expansion and modernization of the road system (roads).

Sub-goal 4: expansion and modernization of the social infrastructure (social infrastructure)

Tasks:
- improving the condition of the healthcare system (healthcare),
- improving the condition of the education system (education),
- increasing the number of cultural and arts facilities (culture and arts),
- increasing the number of sports facilities (sports facilities).

Sub-goal 5: development of the local economy (economy)

Tasks:
- development of tourism (tourism),
- modernization of agriculture (agriculture),

---

[8] The terms to be used later in this paper are put in brackets.

- construction of industrial sites (industrial sites),
- development of residential housing (housing).

With these tasks in mind, three alternative development scenarios were developed:

Scenario 1: multi-functional development through support for housing, services and tourism.

Scenario 2: multi-functional development through support for entrepreneurship and agri-food processing.

Scenario 3: multi-functional development through support for agriculture and diversification of farm activities.

The strategic factors were compared pairwise at each hierarchy level by local experts (councillors[9] of the municipalities under consideration). Based on experts' opinions, local and global priorities were calculated for sub-goals, tasks and development scenarios with the use of AHP and FAHP. The method for calculating priorities is illustrated by the example of tasks within the environmental protection goal for the Międzychód municipality.

Table 3: Verbal and numerical assessment of pairwise comparisons between tasks within the environmental protection sub-goal for the Międzychód municipality

Councillor	**Intensity of importance of the first task over the second one: verbal assessment numerical assessment (triangular fuzzy number)**		
	landfills/ water bodies	**landfills/ alternative energy sources**	**water bodies/ alternative energy sources**
1	weak+ 4 (2, 4, 6)	weak+ 4 (2, 4, 6)	equal importance 1 (1, 1, 1)
2	strong+ 6 (4, 6, 8)	strong 5 (3, 5, 7)	strong+ 6 (4, 6, 8)
3	strong 5 (3, 5, 7)	very strong+ 8 (6, 8, 9)	weak 3 (1, 3, 5)
…	…	…	…
13	strong 5 (3, 5, 7)	strong 5 (3, 5, 7)	absolute 9 (7, 9, 9)
14	equal importance 1 (1, 1, 1)	strong 5 (3, 5, 7)	strong+ 6 (4, 6, 8)
15	strong 5 (3, 5, 7)	strong 5 (3, 5, 7)	equal importance 1 (1, 1, 1)
Geometric mean	1.995 (1.289; 1.995; 2.775)	1.900 (1.189; 1.900; 2.764)	1.256 (0.796; 1.256; 1.786)

+ Means a slightly stronger intensity of importance if the decision maker is unable to choose between two neighboring ratings.

Source: Author's own study based on the results of a survey with the councillors of the Międzychód municipality in 2013.

[9] The number of councilors depends on the number of inhabitants and is set by law (The Act on Commune Self-Government, 1990): 15 councilors for municipalities up to 20,000 inhabitants.

Pairwise comparisons of tasks, presented as linguistic variables, were converted into the corresponding numerical values and averaged using geometric mean (Table 3). These values were arranged into the pairwise comparison matrix in AHP (Table 4), or into the equivalent fuzzy matrix in FAHP (Table 5).

Table 4: Pairwise comparison matrix for tasks under sub-goal 1 and local priority values

i	Tasks	**Pairwise comparison matrix for tasks**			**Normalized pairwise comparison matrix for tasks**			w_i
		landfills	**water bodies**	**alternative energy sources**	**landfills**	**water bodies**	**alternative energy sources**	
1	Landfills	1.000	1.995	1.900	0.493	0.526	0.457	0.492
2	Water bodies	0.501	1.000	1.256	0.247	0.264	0.302	0.271
3	Alternative energy sources	0.526	0.796	1.000	0.259	0.210	0.241	0,237
Total		2,027	3.791	4.156	1.000	1.000	1.000	1.000

$\lambda_{max} = 3.008$, $RI = 0.5245$, $CI = 0.004$, $CR = 0.76\%$.

Source: Author's own calculations based on the results of a survey with the councillors of the Międzychód municipality in 2013.

Table 5: Fuzzy pairwise comparison matrix for tasks under sub-goal 1

i	Tasks	**Fuzzy pairwise comparison matrix for tasks**			$\widetilde{Q}_i$
		landfills	**water bodies**	**alternative energy sources**	
1	Landfills	(1.000; 1.000; 1.000)	(1.289; 1.995; 2.775)	(1.189; 1.900; 2.764)	(0.264; 0.491; 0.865)
2	Water bodies	(0.360; 0.501; 0.776)	(1.000; 1.000; 1.000)	(0.796; 1.256; 1.786)	(0.163; 0.276; 0.471)
3	Alternative energy sources	(0.362; 0.526; 0.841)	(0.560; 0.796; 1.257)	(1.000; 1.000; 1.000)	(0.146; 0.233; 0.410)

Source: Author's own calculations based on the results of a survey with the councillors of the Międzychód municipality in 2013.

The consistency of pairwise comparisons was also checked. In this case, the inconsistency ratio was 0.76% which means that the pairwise comparisons of tasks under sub-goal 1 for the Międzychód municipality were consistent. Next, the pairwise comparison matrix was used to determine the local priority vectors w_i in AHP (Table 4) and FAHP (Table 6). As shown by the results, the most important task under the environmental protection goal was to solve the landfill issue. The local priorities for this task were 0.492 (AHP) and 0.539 (FAHP) which means that its contribution to the first sub-goal was around 50%. The other two tasks had significantly lower values of local priorities (Tables 4 and 6), and therefore were less important for the attainment of the environmental protection sub-goal than the first task.

Table 6: Calculating the local priorities of tasks in FAHP

i	$\tilde{Q}_i$			g	$\tilde{Q}_g$			$V(\tilde{Q}_i \geq \tilde{Q}_g)$	$\min V(\tilde{Q}_i \geq \tilde{Q}_g)$	w_i
	l_i	m_i	u_i		l_g	m_g	u_g			
1	0.264	0.491	0.865	2	0.163	0.276	0.471	1.000	1.000	0.539
1	0.264	0.491	0.865	3	0.146	0.233	0.410	1.000		
2	0.163	0.276	0.471	1	0.264	0.491	0.865	0.492	0.492	0.266
2	0.163	0.276	0.471	3	0.146	0.233	0.410	1.000		
3	0.146	0.233	0.410	1	0.264	0.491	0.865	0.362	0.362	0.195
3	0.146	0.233	0.410	2	0.163	0.276	0.471	0.850		
								total	1.854	1.000

Source: Author's own calculations based on the results of a survey with the councillors of the Międzychód municipality in 2013.

The values of local priorities for other tasks, sub-goals and development scenarios are calculated in a similar way. The next step is to calculate the global priorities. In the case of sub-goals, the local and global priorities are the same. Global priorities for tasks are calculated by multiplying the local priority value of a task by the corresponding global priority value of the sub-goal. When multiplied by the values of global priorities for tasks, the local priorities for scenarios become partial priorities. The global priority is determined only after the partial priorities within a scenario have been added together.

Table 7 shows the values of global priorities for sub-goals and tasks for the Międzychód and Chrzypsko Wielkie municipalities. Note that calculations based on AHP and FAHP showed similar values of global priorities. The sub-goals related to the development of the technical and social infrastructure were important to both municipalities. In the Międzychód municipality, the global priority values for the technical infrastructure calculated using AHP and FAHP were 0.228 and 0.223, respectively. For the social infrastructure, the respective values were 0.220 and 0.215. Similarly, in the Chrzypsko Wielkie municipality, the global priority values calculated using AHP and FAHP for the technical infrastructure were 0.222 and 0.219, respectively, and 0.195 and 0.201 for the social infrastructure. In addition to important sub-goals shared between the municipalities, the urban and rural municipality of Międzychód demonstrated the importance of the environmental protection sub-goal (AHP: 0.211, FAHP: 0.214). The rural municipality of Chrzypsko Wielkie, on the other hand, attached importance to improvements in quality of life (AHP: 0.219, FAHP: 0.216) and local economy development (AHP: 0.223, FAHP: 0.216). All these goals are rated as "medium important" by the municipal councillors (with global priority values at around 0.2). Other goals were slightly less important to the municipalities.

The global priority values calculated with FAHP (Table 7) were used to discuss the importance of tasks. In the urban and rural municipality of Międzychód, one of the most important tasks was solving the landfill issue. The global priority of that task was 0.115 which means that its contribution to the main goal was 11.5%. As regards the rural municipality of Chrzypsko Wielkie, the key tasks were reducing the unemployment (a global priority of 0.127) and extending and upgrading the healthcare infrastructure (0.101).

Table 7: Values of global priorities for sub-goals and tasks for the Międzychód and Chrzypsko Wielkie municipalities

Sub-goals	**Międzychód**		**Chrzypsko Wielkie**		**Tasks**	**Międzychód**		**Chrzypsko Wielkie**	
	AHP	**FAHP**	**AHP**	**FAHP**		**AHP**	**FAHP**	**AHP**	**FAHP**
Environmental protection	**0,211**[a)]	**0,214**	0,142	0,149	landfills	0.104	0.115	0.078	0.084
					water bodies	0.057	0.057	0.037	0.046
					alternative energy sources	0.050	0.042	0.027	0.019
Quality of life	0,160	0,167	**0,219**	**0,216**	unemployment	0.054	0.056	0.107	0.127
					security	0.048	0.055	0.044	0.043
					communication	0.029	0.031	0.038	0.038
					lighting	0.029	0.026	0.030	0.007
Technical infrastructure	**0,228**	**0,223**	**0,222**	**0,219**	water supply network	0.073	0.079	0.077	0.079
					sewage network	0.066	0.069	0.062	0.066
					gas supply network	0.028	0.009	0.020	0.000
					roads	0.062	0.066	0.063	0.074
Social infrastructure	**0,220**	**0,215**	**0,195**	**0,201**	healthcare	0.071	0.076	0.084	0.101
					education	0.070	0.087	0.058	0.077
					culture and arts	0.027	0.000	0.023	0.000
					sport facilities	0.051	0.052	0.030	0.022
Economy	0,181	0,181	**0,223**	**0,216**	tourism	0.035	0.036	0.045	0.043
					agriculture	0.054	0.053	0.079	0.075
					industrial sites	0.047	0.046	0.048	0.047
					housing	0.045	0.046	0.051	0.051

[a)] Priority values for the most important sub-goals are in bold.

The highest values of global priorities for tasks $w_i^g \in \langle 0.09; 0.13\rangle$ are in dark grey; the important tasks $w_i^g \in \langle 0.06; 0.09\rangle$ are in medium grey; the medium important tasks $w_i^g \in \langle 0.03; 0.06\rangle$ are in light grey; tasks $w_i^g \in \langle 0; 0.03\rangle$ which do not have to be urgently implemented are in white.

Source: Author's own calculations based on the results of surveys with the councillors of the Międzychód and Chrzypsko Wielkie municipalities in 2013.

Moreover, the two municipalities attached importance to tasks related to the development and upgrade of the water supply and sewage networks and roads; as well as to tasks involving the upgrade of education facilities. The extension and upgrade of the healthcare infrastructure proved to be important to the

Międzychód municipality. In the rural municipality of Chrzypsko Wielkie two important tasks were identified: solving the landfill issue and developing the agriculture. The global priorities for these tasks fell into the interval (0.06; 0.09⟩.

The use of FAHP allowed also to discover some tasks which do not need to be implemented. In both municipalities, these were increasing the number of cultural and arts facilities. Moreover, the extension of the gas supply network proved not to be necessary in Chrzypsko Wielkie municipality. The global priorities for these tasks were zero.

Table 8: Importance hierarchy of tasks for the Międzychód and Chrzypsko Wielkie municipalities

Sub-goals	**Tasks**	**Międzychód**	**Chrzypsko Wielkie**
		FAHP	**FAHP**
Environmental protection	landfills	+++	++
	water bodies	+	+
	alternative energy sources	+	0
Quality of life	unemployment	+	+++
	security	+	+
	communication	+	+
	lighting	0	0
Technical infrastructure	water supply network	++	++
	sewage network	++	++
	gas supply network	0	0
	roads	++	++
Social infrastructure	healthcare	++	+++
	education	++	++
	culture and arts	0	0
	sport facilities	+	0
Economy	tourism	+	+
	agriculture	+	++
	industrial sites	+	+
	housing	+	+

+++ means the most important task; ++ means an important task; + means a medium important task; 0 means a non-urgent task.

Source: Author's own study based on Table 7 data.

In the Międzychód municipality, the most important task was identified (with a global priority value $w_i^g \in (0.09;\ 0.13\rangle$) together with five important tasks ($w_i^g \in (0.06;\ 0.09\rangle$), ten medium important tasks ($w_i^g \in (0.03;\ 0.06\rangle$) and three non-urgent tasks ($w_i^g \in \langle 0;\ 0.03\rangle$). In the rural municipality of Chrzypsko Wielkie, there were two most important tasks, six important tasks, six medium important tasks, and five non-urgent tasks (Tables 7 and 8).

The proposed importance hierarchy of tasks for the municipalities (Table 8), as supplemented with information on funding sources and operators charged with the performance of specific tasks, may be used when drawing up the development strategies for the municipalities.

Table 9: Global priority values for the development scenarios for the Międzychód and Chrzypsko Wielkie municipalities

Scenarios	Międzychód		Chrzypsko Wielkie	
	AHP	FAHP	AHP	FAHP
Multi-functional development through support for housing, services and tourism	**0.422**	**0.482**	**0.323**	0.309
Multi-functional development through support for entrepreneurship and agri-food processing	0.363	0.396	**0.353**	**0.380**
Multi-functional development through support for the agriculture and diversification of farm activities	0.215	0.122	**0.324**	0.311

Source: Author's own calculations based on the results of surveys with the councillors of the Międzychód and Chrzypsko Wielkie municipalities in 2013.

With these approaches, it was also possible to assess the importance of development scenarios for the municipalities under consideration (Table 9). In the case of the Międzychód municipality, the results of AHP and FAHP analyses were similar. The multi-functional development through support for the housing, services and tourism (with a global priority of 0,422 and 482, respectively) turned out to be the best scenario (Table 9). As regards the Chrzypsko Wielkie rural municipality, the results obtained from AHP did not provide an unequivocal identification of the best development scenario, since the global priorities had similar values (around 0.3). In this case, FAHP proved to be useful in solving this problem as it provided a basis for selecting the development scenario involving multi-functional development through support for entrepreneurship and agri-food processing. The global priority for this scenario was 0.380 (Table 9).

4 Summary

In presented approaches, quantitative methods are combined with qualitative experts' assessments. The use of AHP and FAHP allowed to assess the importance of sub-goals, strategic tasks and development scenarios for municipalities based on experts' opinions. Both methods resulted in similar importance ratings of strategic factors. However, FAHP allowed to "sharpen" the values of global priorities as compared to classic AHP which was particularly evident in the assessment of the development scenarios. The classic AHP has failed to unequivocally identify the best scenario for the Chrzypsko Wielkie municipality. This was only possible using FAHP.

In the Chrzypsko Wielkie municipality, the most important scenario involved the development through support for entrepreneurship and agri-food processing. In the urban and rural municipality of Międzychód, the best scenario involved the development through support for housing, services and tourism.

Moreover, the use of FAHP allowed to eliminate the tasks with the lowest strategic importance (global priority equal to zero). Furthermore, based on task importance assessments by local experts, the importance hierarchy of tasks was established to identify the most important ones (priority tasks), important tasks, medium-important tasks and non-urgent tasks.

The proposed approach to the assessment of importance of strategic factors for a municipality can be used in the planning of socio-economic development of administrative units as the underlying reason for building the development strategy.

References

The Act on Commune Self-Government of 8 March 1990, Dziennik Ustaw (Journal of Laws) 2013, pos. 594, with subsequent (in Polish).

Aguaron J., Moreno-Jiménez J.M. (2003), *The Geometric Consistency Index: Approximated Thresholds*, European Journal of Operational Research, 147(1), 137-145.

Alonso J.A., Lamata M.T. (2006), *Consistency in the Analytic Hierarchy Process: A New Approach*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 14(4), 445-459.

Chang D.-Y. (1996), *Application of the Extent Analysis Method on Fuzzy AHP*, European Journal of Operational Research, 95(3), 649-655.

Csutora R., Buckley J.J. (2001), *Fuzzy Hierarchical Analysis: The Lambda-Max Method*, Fuzzy Sets and Systems, 120(2), 181-195.

Franek J., Kresta A. (2014), *Judgement Scales and Consistency Measure in AHP*, Procedia Economics and Finance, 12, 164-173.

Miller G.A. (1956), *The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, The Psychological Review, 63, 81-97.

Saaty T.L. (1980), *The Analytic Hierarchy Process. Planning, Priority Setting, Resource Allocation*, McGraw-Hill, New York.

Śmigielska K. (2013), *Zastosowanie metody analitycznego procesu hierarchicznego do wyboru scenariuszy rozwoju gmin powiatu międzychodzkiego* [*Application of the Method of Analytic Hierarchy Process for the Selection of Development Scenarios for Municipalities in Międzychód County*], Source material, Uniwersytet Przyrodniczy w Poznaniu (in Polish).

Trzaskalik T. (2014a), *Wielokryterialne wspomaganie decyzji. Przegląd metod i zastosowań* [*Multicriteria Decision Support. Review of Methods and Applications*], Zeszyty Naukowe Politechniki Śląskiej. Seria: Organizacja i Zarządzanie Vol. 1921, 239-263 (in Polish).

Trzaskalik T., ed. (2014b), *Wielokryterialne wspomaganie decyzji* [*Multicriteria Decision Support*], PWE, Warszawa (in Polish).

Wang J.-W., Cheng C.-H., Kun-Cheng H. (2009), *Fuzzy Hierarchical TOPSIS for Supplier Selection*, Applied Soft Computing, 9(1), 377-386.

M U L T I P L E   C R I T E R I A   D E C I S I O N   M A K I N G
Vol. 12 2017

David M. Ramsey*

THE USE OF INITIAL FILTERS TO DIRECT SEARCH IN DECISION PROCESSES

DOI: 10.22367/mcdm.2017.12.10

Abstract

When looking for a valuable resource, many people use information from the Internet as a way of choosing a small number of offers to investigate in more detail. This paper considers strategies based on the filtering of initial information. A new model is presented according to which the goal of the decision maker is to maximise her expected reward from search taking into account the search costs. The effectiveness of strategies based on filtering is compared to sequential search based on a threshold and exhaustive search of a chosen number of items.

Keywords: filtering, job search problem, multiple signals, sequential search.

1 Introduction

When searching for a valuable good (e.g. a car or house), individuals very often use a two-stage process. In the first stage, the Internet or a specialist magazine may be used in order to find a number of offers that seem promising. This stage is normally characterised by the ability to compare general information regarding a large number of offers at relatively low cost. This stage may be thought of as wide, shallow search. On the basis of this initial information, the searcher may choose a number of offers to investigate more closely. It is assumed that the search costs for obtaining additional information at this stage are relatively large. Hence, the second stage may be called narrow, deep search. During (or after) this stage, the searcher may decide to purchase one of the offers and thus stop searching, or to return to the first stage of searching. Hence, the search process can be charac-

* Wrocław University of Science and Technology, Faculty of Computer Science and Management, Department of Operations Research, Wrocław, Poland, e-mail: david.ramsey@pwr.edu.pl.

terised as being partially parallel, since a number of offers can be considered at the same time. It is also partially sequential, since if the searcher returns to the first stage of search, then only new offers will be considered.

MacQueen and Miller (1960) presented a model of sequential search. A decision maker observes a (theoretically unlimited) sequence of offers whose values come from a known distribution and must at each stage decide whether to accept or reject an offer. A fixed cost is paid for observing each offer. Once an offer has been accepted, then the decision maker ceases searching. The optimal strategy for this problem is to accept the first offer whose value exceeds the expected reward from search, i.e. a threshold strategy. Various variations of such a problem have also been considered in the economics literature, in terms of a consumer searching for a job or product, and in the biological literature, particularly as models of mate choice in which just females are choosy. Stigler (1961) considered a similar model of consumer search. He assumed that the strategy of a consumer, the so called "best-of-n rule", is defined by the number of offers to observe, n. The consumer accepts the best of the n offers observed. The optimal strategy, which observes n^* offers, satisfies the condition that n^* is the largest integer n satisfying the condition that the expected marginal gain from observing the n-th offer (calculated before any offers are observed) is greater than the cost of observing the n-th offer. Hence, Stigler assumes that a consumer can return to a previously observed offer. However, given that the number of offers is unlimited and search costs are proportional to the number of objects seen, the threshold rule presented by MacQueen and Miller (1960) indicates that the recall of previous offers is of no value. Janetos (1980) compares various types of search rules, including threshold rules and best-of-n rules, in the context of mate choice. Based on his assumptions, the number of prospective mates was essentially fixed and search costs were not specifically modelled, Janetos concluded that best-of-n rules are the most effective. Real (1990) incorporates search costs into his model of mate choice and shows that when search costs are proportional to the number of prospective mates seen, then threshold rules are optimal. Hutchinson and Hałupka (2004) consider a model in which the population of males is patchy (i.e. search costs are not proportional to the number of males seen). They conclude that best-of-n type rules work best when males group together in so-called leks (i.e. recall of recently seen prospective partners is possible and cheap) and there is little spatial variation in the values of males. When males are more dispersed (i.e. recall of previously seen prospective partners is expensive or impossible) and/or there is spatial variation in the values of mates, then threshold rules are more successful.

The articles considered above assume that all the information regarding the value of an offer is observed at a single moment. MacQueen (1964) extended this approach to model sequential search in which the decision maker can gain additional information about an offer, which incurs additional inspection costs. The decision maker first observes a quantitative signal of the value of an object and on the basis of this observation chooses one of the following three options: a) immediately accept the offer and stop searching, b) gain additional information, in the form of a second signal, about the offer, c) reject the offer and continue searching. In case b), after observing the additional information, the decision maker must decide whether to accept or reject the offer. When the joint distribution of the two signals is known and the search costs are linear in both the number of offers seen and the number of times that the second signal is observed, then the optimal strategy is defined by a vector of three thresholds (t_1, t_2, t_3). If the quantitative measure of the first signal is less than t_1, then an offer should be immediately rejected. If this measure is greater than t_2, then the offer should be immediately accepted. Otherwise, additional information should be gathered about the offer. After the second signal has been observed, then the offer should be accepted if and only if the value of the object based on the two signals is at least t_3. Ramsey (2015) derives the specific form of this optimal policy when the two signals come from a joint normal distribution and the value of an offer is equal to a linear combination of the two signals. This approach is based on the concept of multi-attribute linear utility (MLU, see Keeney and Raiffa, 1993), which is one approach sometimes used to solve multi-criteria decision problems. Lim et al. (2006) extend this model to one where multiple signals describing the value of an offer are available. It is assumed that these signals must be observed in a particular order and that the expected value of an offer given the value of a signal is increasing in the quantitative measure of this signal. They define a dynamic programming procedure that derives the optimal policy for such a problem. Wiegmann et al. (2010) consider a similar problem within the framework of a mate choice problem. Bearden and Connolly (2007a, 2007b) consider an approach to such problems based on the concept of satisficing (see Simon, 1995). Based on such a strategy, the second signal is observed only when the quantitative measure of the first signal exceeds the appropriate threshold. Given that the second signal is observed, then the offer is accepted if and only if the quantitative measure of this signal exceeds the specified threshold. They show that such strategies can be close to optimal and on the basis of experiments find that individuals who are restricted to using satisficing strategies are on average as successful as individuals who are allowed to use strategies of the same form as the optimal strategy. Chun (2015) uses a similar approach, but assumes that the most important trait is optimised subject to satisficing other conditions.

Hogarth and Karelaia (2005) consider a problem in which a number of binary signals can be observed for each offer and offers can be observed in parallel. In accordance with the concept of satisficing, these signals can be interpreted as stating whether a given condition is satisfied or not. It is assumed that these conditions can be ranked according to importance. One approach to selecting an offer is the DEBA Procedure (Deterministic Elimination by Aspects). The decision maker starts with the most important condition and eliminates those offers that do not satisfy this condition. The decision maker then passes on to successively less important criteria and at each stage eliminates those offers that do not satisfy the present criterion. If only one offer is left, then that offer is chosen. If none of several remaining offers satisfy the present condition, or all of the conditions have been considered and several offers still remain, then an offer is chosen at random from those remaining. The effectiveness of such a procedure is compared to the so-called Equal Weighting Procedure (EW), according to which all the signals are observed for each offer and the offer that satisfies the most conditions is accepted (or a random choice made between the offers satisfying the most criteria). It is shown that when preferences are non-compensatory, i.e. the weight of a given condition is greater than the sum of the weights of less important criteria, then the DEBA procedure outperforms the EW procedure. Baucells et al. (2008) investigates the effectiveness of the DEBA procedure under various assumptions. It should be noted that in problems of this type the decision maker can choose in which order to observe the signals, unlike in the problem considered in this paper, where it is assumed that there is a natural order in which the signals must be observed. Also, these models do not explicitly model the costs of searching or making decisions. In the context of mate choice, Fawcett and Johnstone (2003) consider a similar problem where two binary traits are associated with the value of a prospective mate, but they define explicit costs for observing a trait. They derive conditions for the optimality of i) random mating, ii) mating based on one signal and iii) mating based on both signals. In the case that both signals should be considered, they derive a condition which determines which signal should be observed first.

Analytis et al. (2014) consider a two-stage search process, in which the information obtained in the first stage is used to decide in which order the offers should be investigated in the second round. Hence, in the first round comparison is made in parallel. The second round involves sequential inspection of the offers, starting with the offer that was adjudged to be the best in round one. At this stage, an offer is accepted and search ceases when the expected value of the offer given the second signal is greater than the expected reward from continued search (i.e. a threshold rule is used). This approach is developed in Analytis et al.

(2015), where a decision maker is searching for a good whose utility depends both on his/her preferences and (positively) on the popularity of that good (this might be appropriate for modelling the utility of e.g. a book or CD). Since Internet suppliers often order goods according to popularity, buyers use this as an initial filter in a sequential search procedure. Chhabra et al. (2014) look at a similar problem from the point of view of an information provider who wishes to set prices for services.

This article is intended as a starting point for building a model of searching for a valuable good which has unique (or near unique) characteristics, e.g. a house or second hand car. In the first round of search, information can be gained on a relatively large number of offers in parallel at low cost. The search costs incurred in the second round are assumed to be greater. Also, it is assumed that the information obtained in the first stage is of a more quantitative nature than the information gained in the second stage. For example, the description of a flat in the Internet will give specific information about the size, price and location of the flat, which can be evaluated in numerical terms. Visiting a flat will give additional information, which will tend to be of a more qualitative nature. Hence, the model assumes that in the first stage of search a quantitative measure of the value of an offer will be observed and after the second stage the searcher can only rank the offers.

Another of the goals of the article is to highlight conditions under which a strategy based on filtering can be optimal. Under such a strategy, on the basis of the first stage, the decision maker should choose a relatively small number of promising offers to observe more closely in the second round, before making a final decision. In particular, a comparison of the article of MacQueen (1964) with the research of Hutchinson and Hałupka (2004) is particularly useful in this regard. Sequential search using a threshold strategy, rather than a filtering strategy, will be optimal when the search costs are proportional to the number of offers observed and the searcher perfectly observes quantitative signals, as well as having perfect information about the distribution from which signals come (and their relation to the value of an offer). Hutchinson and Hałupka (2004) note that when the search costs are not linear and recall of previously seen mates is possible (this is the case when offers are observed in parallel), then best-of-*n* type rules may work very well. This is also true when decision makers do not have perfect information regarding the distribution of the value of offers. Also best-of-*n* type rules will work well when signals cannot be observed precisely, but can only be ranked according to attractiveness (for example, it is impossible to say anything about the attractiveness of the first offer observed, since it cannot be compared with anything). It follows that filtering strategies should be effective when one or more of the following conditions are satisfied:

1. In the first round a number of offers can be observed in parallel and the average search cost per offer is decreasing in the number of observations.
2. The decision maker does not have perfect information regarding the distribution of the signals or their relation to the value of offers.
3. Some of the signals cannot be precisely measured, but are comparable.

The model considered here assumes that the second signal cannot be precisely measured. However, search costs are assumed to be proportional to the number of offers investigated and it is implicitly assumed that the decision maker knows the distribution of the signals. Intuitively, if the amount of information contained in the second signal is sufficiently large, then the optimal strategy should be a filtering strategy.

Section 2 presents the model and describes how the optimal strategy is derived. Three types of strategy which are particularly important in this framework are highlighted: i) sequential search using a threshold based purely on the first signal, ii) full investigation of a number n of observations, iii) filtering strategies. Section 3 presents an example where the two signals have exponential distributions. This is used to illustrate the derivation of the optimal strategy. Section 4 presents the results and, in particular, concentrates on the form of the optimal policy according to the relative amount of information contained in the second signal. Conclusions and some directions for future research are given in Section 5.

2 A model with incomplete information

This section considers a new model of two-stage search in which the signal observed in the first round of search can be summarised using a quantitative variable, X_1, whereas the overall impression of an offer after the second round can only be ranked in comparison to the other offers observed within a batch. However, it is assumed that the overall value of the object is a quantitative value that comes from some distribution conditional on X_1. The cost of inspecting an offer in round i is defined to be c_i, $i = 1,2$.

The strategy of a searcher is defined by a vector of two variables (n, t), where n is the number of objects observed in a batch and t is the threshold used in the first round. Offer j, $1 \leq j \leq n$ passes through to round two if and only if $X_{1,j} \geq t$, where $X_{1,j}$ is the j-th observation of the variable X_1. If no offer from a batch exceeds this threshold, then the searcher observes another batch. If exactly one offer from a batch exceeds this threshold, then this offer is automatically accepted without incurring any search costs in the second round. If l offers from a batch pass through to round two, where $l > 1$, then these offers are compared and the appropriate search costs are incurred, i.e. $c_2 l$. Note that during the second round it suffices that the searcher remembers the best offer seen so far and compares this to each new offer. After observing all l offers, the searcher accepts the most highly ranked offer.

From the description of such strategies, a strategy with a batch size of one is a sequential search policy based purely on the first signal and defined by the appropriate threshold. Assuming that X_1 is a positive random variable, it follows that for $n > 1$, an individual following the strategy (n, 0) will fully inspect all n objects in a batch and accept the best one seen. For these reasons, a strategy (n, t) will be called a filtering strategy when $n > 1$ and $t > 0$.

The goal of the searcher is to maximise the expected reward from search, which is assumed to be the value of the offer accepted minus the search costs incurred. Define $W(n, t)$ to be the random variable denoting the reward of a searcher using the strategy (n, t). Let $L \equiv L(n, t)$ be the number of offers that pass through to the second round. It follows that $L \sim \text{Bin}(n, p)$, where $p = P(X_1 > t)$. Conditioning on the number of offers that pass through to the second round, from the form of the strategy, we obtain:

$$E[W(n,t)|L = 0] = E[W(n,t)] - c_1 n \quad (1)$$

$$E[W(n,t)|L = 1] = E[V|X_1 > t] - c_1 n \quad (2)$$

$$E\left[\max_{1 \le j \le l}\{V_j|X_{1,j} > t\}\right] - c_1 n - c_2 l, \text{ for } 2 \le l \le n, \quad (3)$$

where V_j is the overall value of the j-th offer seen in round 2. It follows that the value function, $E[W(n, t)]$ may be derived using the adaptation of the law of total probability to expected values. A two-step procedure is used to find the optimal strategy. Firstly, the optimal threshold for a given batch size is found and then we optimize over the set of batch sizes. Note that we may obtain an estimate of how large the optimal batch size is by considering the expected number of items that should be seen under a policy based purely on the first signal. This is considered in the next section on the basis of an example. The optimal threshold for a given batch size is found using a program written in R using the *Optimx* package (see Kelley, 1999). Since the value function can have local optima, these functions were graphed in order to ensure that the global maximum was selected.

3 Example

It is assumed that the first signal, X_1, has an exponential distribution with mean 1 and the second signal, X_2, has an exponential distribution with mean μ, independently of the first signal. The value of an offer is given by $V = X_1 + X_2$. The value of the first signal is observed precisely, while the values of the offers can only be ranked. Since the standard deviation for an exponential distribution is equal to the mean, the proportion of the information about the value of an offer described by the value of the second signal (if it could be precisely observed) may be defined by $q = \frac{\mu}{1+\mu}$. The cost of observing the first signal is $c_1 = 0.01$ and the cost of closer inspection is $c_2 = 0.1\mu$. Hence, in the second round the cost of search relative to the amount of information gained is 10 times greater than in the first round.

In order to estimate the optimal batch size and the expected number of offers from a batch that pass through to round two when a filtering policy is optimal, we consider two auxiliary search problems. The first problem assumes that the searcher uses a sequential search policy based purely on observations of the first signal (i.e. uses a batch size of one). The second problem assumes that the searcher uses a policy of the form: compare k values of the second trait and choose the best one. From the linearity of both the search costs and the value of an offer as a function of X_1 and X_2, as well as the independence of these two variables, it is expected that the optimal batch size should be close to the expected number of offers seen under the optimal strategy in the first auxiliary problem and the expected number of offers passing through to the second round should be close to the optimal choice of k in the second problem.

In the first auxiliary problem, an offer should be accepted only when its value is greater than the future expected reward from search (ignoring any costs already incurred), i.e.:

$$\begin{aligned} t = E[\max\{X_1, t\}] - c_1 = E[X_1 | X_1 > t]P(X_1 > t) + tP(X_1 < t) - c_1 = \\ = (1+t)e^{-t} + t(1 - e^{-t}) - c_1. \end{aligned} \tag{4}$$

The second line follows from the lack of memory property of the exponential distribution. Hence, the optimal threshold satisfies $e^{-t} = c_1 \Rightarrow t = \ln(c_1^{-1})$. Thus the probability of accepting an offer under the optimal policy is $P[X_1 > \ln(c_1^{-1})] = c_1$ and the expected number of offers observed before one is accepted is c_1^{-1}.

In the second problem, the strategy is defined by the number of offers to be inspected. After observing these offers, the searcher chooses the best one. Note that it is better to observe k objects rather than just $k-1$ if and only if the expected increase in the value of the best object seen (based entirely on X_2) is at least as great as the inspection costs. The probability that the k-th offer is the best is $1/k$. Given that the k-th offer is the best, then the amount by which the value of this offer exceeds the value of the previously best offer has an exponential distribution with mean μ (again from the lack of memory property of the exponential distribution). It follows that the optimal choice of k is given by the largest value of k that satisfies $\frac{\mu}{k} \geq c_2 \Rightarrow k \leq \frac{\mu}{c_2}$.

Using these two auxiliary problems, under a strategy involving filtering, the size of a batch should be around 100 and the expected number of objects seen in the second round should be close to 10.

Now consider the search problem with two-step inspection. Under the filtering strategy (n,t), the probability of an offer passing through to the second round is $P(X_1 > t) = e^{-t}$. Hence, the number of offers observed in round 2, L, has a binomial distribution with parameters n and e^{-t}.

Let $X_{i,j}$ denote the numerical measure of the i-th signal for the j-th offer to be seen in round 2. Using the lack of memory property of the exponential distribution, the particular forms of Equations (2) and (3) giving the expected reward from search conditional on the number of offers observed in the second round are given by:

$$E[W(n,t)|L=1] = 1 + t + \mu - c_1 n \quad (5)$$

$$E[W(n,t)|L=l] = t + \max_{1\le j\le l}[X_{1,j} + X_{2,j}] - c_1 n - c_2 l, \text{ for } 2 \le l \le n. \quad (6)$$

From the form of Equation (6), it is clear that in order to solve this problem we should derive the distribution of $\max_{1\le j\le l}[X_{1,j} + X_{2,j}]$. The density function of the value of an offer, $V = X_1 + X_2$, can be derived from the convolution of X_1 and X_2, i.e.:

$$f_V(v) = \int_0^v f_{X_1}(x) f_{X_2}(v-x)dx. \quad (7)$$

This gives:

$$f_V(v) = \begin{cases} \dfrac{\exp\left(-\frac{v}{\mu}\right) - \exp(-v)}{\mu - 1}, & \mu \neq 1 \\ ve^{-v}, & \mu = 1. \end{cases} \quad (8)$$

Since the values of the offers are independent, for $\mu \neq 1$ the distribution function of the maximum value of a set of l offers, U_l, is given by:

$$F_{U_l}(v) = P(U_l \le v) = P(V \le v)^l = \left[1 - \frac{\mu\exp\left(-\frac{v}{\mu}\right) - \exp(-v)}{\mu - 1}\right]^l. \quad (9)$$

Since U_l is a positive random variable, it follows that:

$$E(U_l) = \int_0^\infty [1 - F_{U_l}(v)]dv = \int_0^\infty 1 - \left[1 - \frac{\mu\exp\left(-\frac{v}{\mu}\right) - \exp(-v)}{\mu - 1}\right]^l dv. \quad (10)$$

Expanding the integrand in this expression, it follows that $E(U_l) = \sum_{k=1}^{l} I_k$, where:

$$I_k = (-1)^{k+1} \int_0^\infty \binom{l}{k} \left[\frac{\mu \exp\left(-\frac{v}{\mu}\right) - \exp(-v)}{\mu - 1} \right]^k dv. \tag{11}$$

These integrals can be derived by expanding in a similar manner. It follows that:

$$I_k = \frac{1}{(\mu-1)^k} \sum_{i=0}^{k} (-1)^i \binom{k}{i} \frac{\mu^{k+1-i}}{k + i(\mu - 1)}. \tag{12}$$

Applying the analogue of the law of total probability to the expected reward from search, we have:

$$E[W(n,t)] = \sum_{l=0}^{n} E[W(n,t)|L = l]P(L = l). \tag{13}$$

Using Equations (1), (5) and (6) and rearranging, we obtain:

$$E[W(n,t)] = \frac{1}{1-(1-e^{-t})^n} [ne^{-t}(1-e^{-t})^{n-1}(1+t+\mu) +$$

$$+ \sum_{l=2}^{n} \left\{ (E[U_l] - c_2 l) \binom{n}{l} e^{-tl} (1-e^{-t})^{n-l} \right\} - c_1 n. \tag{14}$$

For a given n, this function was optimised using the Optimx package in R and then the value function was globally maximised by allowing n to vary.

4 Results

The results are illustrated in Figures 1 to 6. Figures 1 to 3 depict how the form of the optimal policy changes as the proportion of information contained in the second signal changes (Figure 1 illustrates the optimal batch size, Figure 2 presents the corresponding optimal threshold and Figure 3 illustrates the expected number of objects inspected in the second round). When the proportion of information contained in the second signal is less than about 65%, then the optimal batch size is equal to one. This corresponds to the optimal sequential search policy based purely on the first signal. Figure 2 indicates that in this region a very high value of the first signal is necessary (and in this case sufficient) for an offer to be accepted. It might be surprising that the second signal must contain so much information in order to be considered, but this seems to result from the following three factors: a) the order in which the signals are observed promotes the importance of the first signal in the decision process, b) the first signal can be measured quantitatively, whereas the second signal is qualitative (at least in terms of assessment by the decision maker), c) the costs of inspecting the second signal are relatively more expensive.

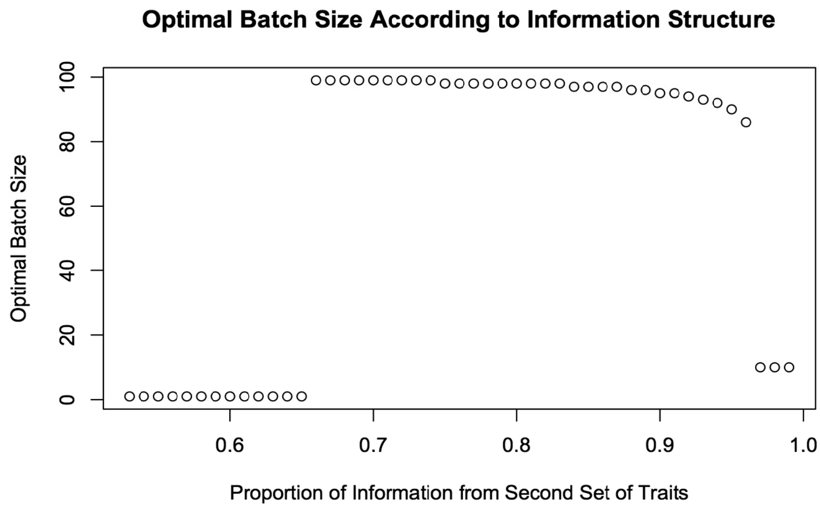


Figure 1. The optimal batch size according to the proportion of information contained in the second signal

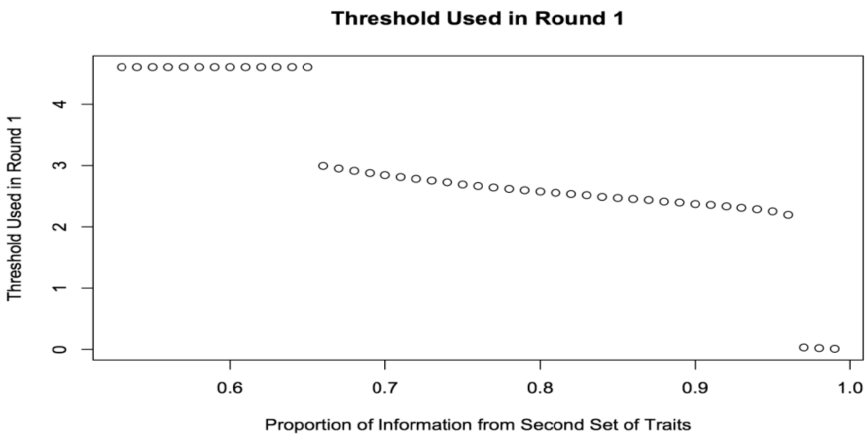


Figure 2. The optimal threshold (corresponding to the optimal batch size) according to the proportion of information contained in the second signal

When the proportion of information contained in the second signal is relatively high (between around 65% and 97%), then the batch size is close to 100. The threshold used in the first round is now quite high, rather than very high. This indicates that a filtering strategy is being used, such that the decision maker uses the first signal as a cheap way of restricting costly inspection only to offers that seem promising. As the second signal becomes more informative, in comparison to the first signal, the optimal batch size falls slowly. On the other hand, the expected number of offers passing through to round two steadily rises (see Figure 3), since the threshold used in the first round decreases. Hence, as one would expect, when the second signal becomes more informative, the searcher will place more importance on observing this signal, compared to observing the first signal. In all cases, the expected number of offers observed in round two is around five or more, which means that the probability of not accepting an offer from a batch is very small. Hence, the variance in the search costs incurred will be relatively small. This is important in the case of risk-averse searchers, who are willing to accept a lower expected reward than the optimal one, if the variance of this reward can be sufficiently reduced.

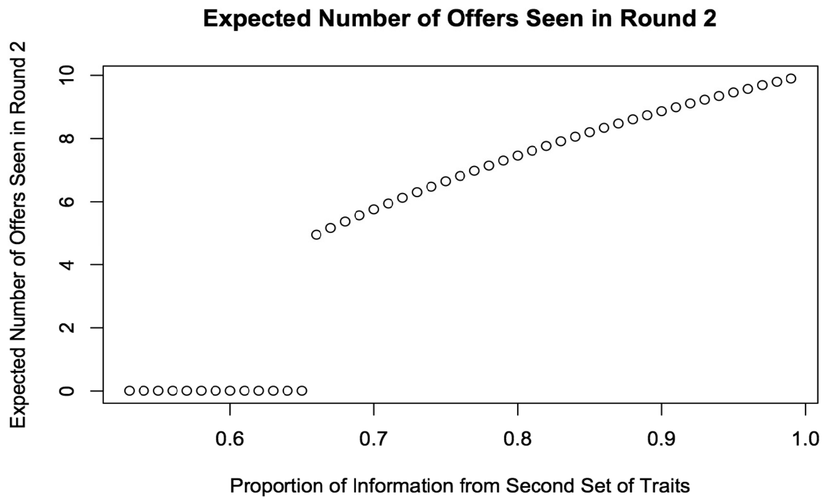


Figure 3. The expected number of offers seen in round 2 according to the proportion of information contained in the second signal

When the proportion of information contained in the second signal is very high (around 97% or more), then the optimal batch size is ten. The threshold used in round one is very close to zero, i.e. almost all the offers seen pass through to round two. This indicates that the first signal is of too little importance to be used as an initial filter to direct search. In this case, an optimally behaving searcher will essentially fully inspect the appropriate number of offers as indicated by the optimal strategy based purely on the second signal and accept the best offer seen.

Figure 4 illustrates the relative values of the expected rewards obtained under the three types of strategy described above: i) the optimal strategy based on filtering, ii) the optimal threshold strategy based purely on the first signal, iii) the strategy based on fully inspecting a batch of ten offers. The expected reward obtained under the optimal filtering policy is scaled to be equal to one. For the problem investigated here, filtering can lead to clear benefits when the second signal contains around 80% of the information about the value of an offer. In this case, the first signal contains enough information to direct search at relatively low cost, while the amount of information contained in the second signal ensures that it should be considered in decision making (even when no quantitative measure of the value of this signal is observable).

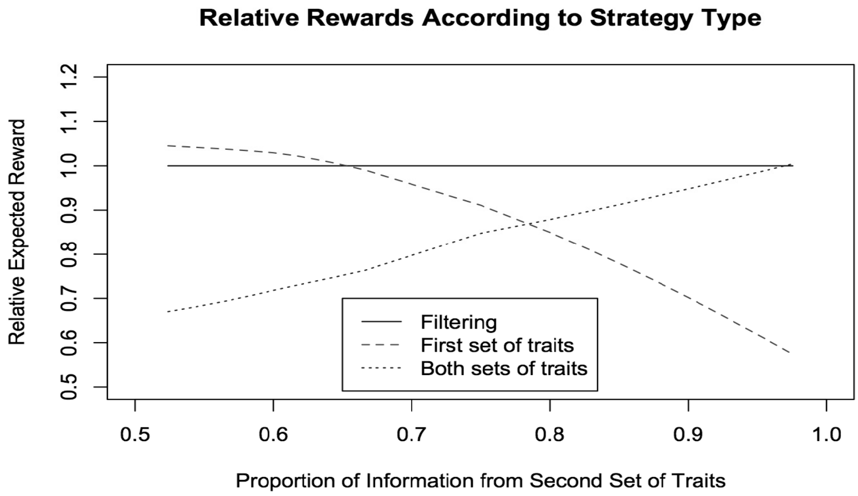


Figure 4. Relative rewards gained according to the type of strategy used and proportion of information contained in the second signal

Figure 5 illustrates how the choice of batch size affects the expected reward from search when the second signal contains 80% of the information. Note that the threshold used is optimally adapted to the batch size. As long as the chosen batch size is not radically different from the optimal batch size, then there is very little effect on the expected payoff. In fact, for the example considered, the expected reward obtained by using any batch size of between 60 and 149 and the corresponding optimal threshold is within 1% of the optimal expected reward. This is attained by using a batch size of 98.

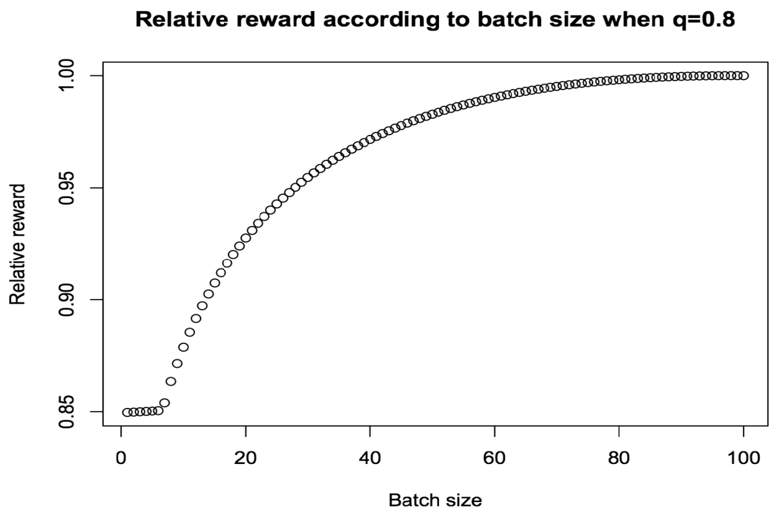


Figure 5. Effect of the batch size on the expected reward when 80% of the information is contained in the second signal

Figure 6 illustrates how the expected number of offers observed in the second round depends on the batch size used. It can be seen that as long as the batch size is not very small in relation to the expected number of offers seen under the optimal strategy based on the first signal (here 100), then the expected number of offers observed in the second round is almost constant. This suggests that a promising line of research would be to consider filtering strategies in which a fixed number of offers from a batch pass through to the second round. The author has some results for strategies of this type, but the solution of such problems can be combinatorically very complex. Hence, such strategies will be considered in a future paper.

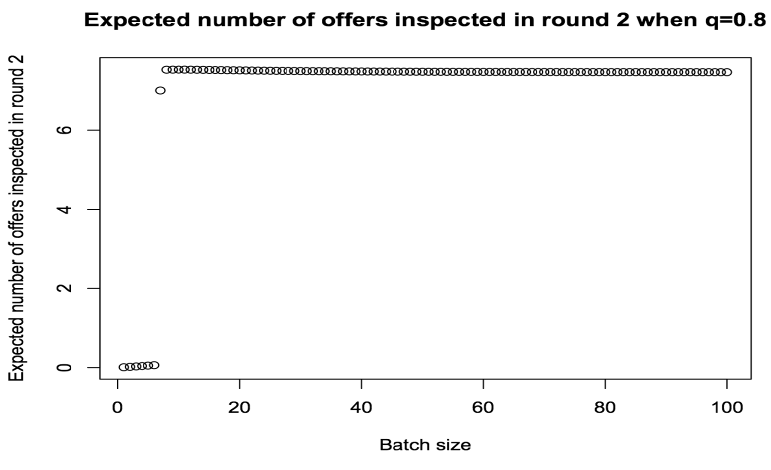


Figure 6. The expected number of offers seen in round 2 according to the batch size when 80% of the information is contained in the second signal

5 Conclusions

This paper has considered a model of searching for a valuable good in which additional information can be gained on chosen offers at some cost, before a final decision is made. In particular, this article has considered the concept of filtering strategies. These are strategies under which initial information on a number of offers is used to choose a smaller number of promising offers to investigate in more detail in the second stage. Such strategies may be beneficial when a number of offers can be observed in parallel and searchers do not have full information about the distribution of the value of offers and/or cannot precisely measure signals associated with the value of an offer. According to the model presented here, the initial filtering of offers is based on a quantitatively measurable initial signal that is cheap to observe. A set of offers which seem promising to the decision maker are then more closely observed on the basis of a second qualitative signal, such that the values of the offers can be ranked. The observation costs in the second round are relatively high compared to those incurred in the first round. These assumptions seem reasonable for the problem of searching for a flat using the Internet to gain initial information about offers before deciding which flats to visit. It is assumed that search costs are proportional to both the number of offers observed and the number of offers that are further investigated in the second round. On the basis of an example, it is shown that such a filtering process is optimal when the amount of information contained in the second signal is sufficiently high, but the first signal also contains some information. When the first signal contains very little information, the optimal strategy is very similar to a ”best-of-n” strategy, i.e. n offers are fully investigated and the best one is accepted. When the first signal contains a relatively large amount of information, then the optimal strategy is to use a threshold policy based on the first signal.

The assumption regarding the proportionality of search costs to the number of offers investigated is a simplification. As argued above, relaxing this assumption by considering the benefits of being able to observe offers in parallel would increase the benefits of using a filtering strategy. Another obvious adaptation of the model would be to assume that neither signal can be measured precisely, but offers can be compared at each stage. In this case, one should consider strategies under which the best k offers according to the first signal should be investigated in the second round. The author has some results regarding such a model. However, the solution of an example analogous to the one presented here is combinatorically complex and will be considered in future research. Another way of making the model more realistic would be to assume that the searcher does not choose the batch size, but this results from the rate at which new offers appear on the market and on the frequency with which an individual observes offers.

In terms of multiple criteria decision making, one interesting characteristic of filtering strategies is that they give very tight control over the search costs incurred while ensuring a valuable offer is obtained. It thus seems reasonable that filtering strategies give a high expected reward from search while keeping the variance of this reward low. This is an important feature, since individuals tend to be risk averse, and should be considered in future research.

One obvious practical problem with the model presented here is that the optimal solution depends on the distribution of the signals and the values of offers. The precise form of the optimal solution will depend on the form of these underlying distributions. In particular, the total variance of the value of an offer and the residual variance of the value of an offer given the first signal are very important in determining how many offers should be considered in each round. These variances depend on both the covariance matrix for the signals, and the relation of the value of an offer to the signals observed. One further avenue for future research would be to investigate heuristic methods of choosing an appropriate number of offers to compare based on the variation in the signals already seen. Also, it would be useful to investigate how robust such an approach would be to the precise form of the joint distribution of signals. For example, Connolly and Wholey (1988) note that this is a difficult problem to solve in practice, even when decision makers possess very good information about the distribution of signals and their relation to the value of an offer. However, it might be the case that an adaptation of the heuristic strategy proposed by Janetos (1980), i.e. "choose five offers for further investigation and accept the best" might be a fairly robust strategy to implement in the second round of such a search procedure.

References

Analytis P.P., Kothiyal A., Katsikopoulos K. (2014), *Multi-attribute Utility Models as Cognitive Search Engines*, Judgment and Decision Making, 95, 403-419.

Analytis P.P., Stojic H., Moussaïd M. (2015), *The Collective Dynamics of Sequential Search in Markets for Cultural Products* [in:] Analytis P.P. (2015), *Psychological Process Models and Aggregate Behavior*, 81-124.

Baucells M., Carrasco J.A., Hogarth R.M. (2008), *Cumulative Dominance and Heuristic Performance in Binary Multiattribute Choice*, Operations Research, 56(5), 1289-1304.

Bearden J.N., Connolly T. (2007a), *Optimal Satisficing*, Management Science, 1(1), 1-14.

Bearden J.N., Connolly T. (2007b), *Multi-attribute Sequential Search*, Organizational Behavior and Human Decision Processes, 103(1), 147-158.

Chhabra M., Das S., Sarne D. (2014), *Expert-mediated Sequential Search*, European Journal of Operational Research, 234(3), 861-873.

Chun Y.H. (2015), *Multi-attribute Sequential Decision Problem with Optimizing and Satisficing Attributes*, European Journal of Operational Research, 243(1), 224-232.

Connolly T., Wholey D.R. (1988), *Information Mispurchase in Judgment Tasks: A Task-driven Causal Mechanism*, Organizational Behavior and Human Decision Processes, 42(1), 75-87.

Fawcett T.W., Johnstone R.A. (2003), *Optimal Assessment of Multiple Cues*, Proceedings of the Royal Society of London B: Biological Sciences, 270, 1637-1643.

Hogarth R.M., Karelaia N. (2005), *Simple Models for Multiattribute Choice with Many Alternatives: When It Does and Does Not Pay to Face Trade-offs with Binary Attributes*, Management Science, 51(12), 1860-1872.

Hutchinson J.M.C., Hałupka K. (2004), *Mate Choice When Males Are in Patches: Optimal Strategies and Good Rules of Thumb*, Journal of Theoretical Biology, 231(1), 129-151.

Janetos A.C. (1980), *Strategies of Female Mate Choice: A Theoretical Analysis*, Behavioral Ecology and Sociobiology, 7(2), 107-112.

Keeney R.L., Raiffa H. (1993), *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Cambridge University Press, Cambridge, UK.

Kelley C.T. (1999), *Iterative Methods for Optimization*, Society for Industrial and Applied Mathematics.

Lim C., Bearden J.N., Smith J.C. (2006), *Sequential Search with Multiattribute Options*, Decision Analysis, 3(1), 3-15.

MacQueen J.B. (1964), *Optimal Policies for a Class of Search and Evaluation Problems*, Management Science, 10(4), 746-759.

MacQueen J., Miller Jr R.G. (1960), *Optimal Persistence Policies*, Operations Research, 8(3), 362-380.

Ramsey D.M. (2015), *On a Sequential Decision Process Where Offers Are Described by Two Traits*, Multiple Criteria Decision Making, 10, 141-154.

Real L. (1990), *Search Theory and Mate Choice. I. Models of Single-sex Discrimination*, The American Naturalist, 136(3), 376-405.

Simon H.A. (1995), *A Behavioral Model of Rational Choice*, The Quarterly Journal of Economics, 69(1), 99-118.

Stigler G.J. (1961), *The Economics of Information*, The Journal of Political Economy, 69(3), 213-225.

Wiegmann D.D., Weinersmith K.L., Seubert S.M. (2010), *Multi-attribute Mate Choice Decisions and Uncertainty in the Decision Process: A Generalized Sequential Search Strategy*, Journal of Mathematical Biology, 60(4), 543-572.

Part II

Regularly contributed papers

MULTIPLE CRITERIA DECISION MAKING
Vol. 12 2017

Andrzej Kobryń*

DEMATEL AS A WEIGHTING METHOD IN MULTI-CRITERIA DECISION ANALYSIS

DOI: 10.22367/mcdm.2017.12.11

Abstract

Modelling of a decision-maker's preferences in multi-criteria decision analysis is performed using weights that reflect the relative importance of the given decision criteria. The determination of accurate values of the weights is therefore of considerable importance. Numerous means and methods are used for this purpose, such as: the entropy method (i.e., the method of objective weights), the Simos method, the SWARA method, the ANP or AHP methods, and many others. This paper analyses the DEMATEL method, frequently used to identify cause-and-effect relationships. Nowadays, it is often used in multi-criteria decision analyses. In the opinion of some authors, DEMATEL may be useful also to determine the weights of criteria. However, the approach presented by these authors has certain drawbacks. The present paper proposes a different approach to the weighting procedure using DEMATEL. Using numerical examples, it also presents weights determined by this method and compared to those obtained using the AHP method.

Keywords: DEMATEL method, multi-criteria decision analysis, weights of criteria.

1 Introduction

Weighting of criteria plays a key role in solving multi-criteria decision problems. As is known, the preferences of the decision-maker related to individual criteria have the form of weights expressing the relative significance of the criteria. In certain circumstances, it is possible to determine weights using the entropy method, as presented, i.a., in papers by Shannon and Weaver (1963) or Ignasiak (2001). The entropy method constitutes a reasonably objective means of defining

* Bialystok University of Technology, Faculty of Civil and Environmental Engineering, Bialystok, Poland, e-mail: a.kobryn@pb.edu.pl.

weights, which allows for estimation of the importance of the analysed criteria on the basis of ratings discrepancies of analysed variants with respect to each criterion. Therefore, such methods of weighting are classified in the literature as objective weighting methods (Deng et al., 2000). In general, however, we have to deal with the situation when the some of the criteria are preferred by the decision-maker more than others. Therefore, the entropy method is of little use for criteria weighting.

In such cases, the most profitable method seems that suggested by Simos (1990). It involves determining criteria weights based on the opinion of the group of people using two card sets of the same size (Simos, 1990; Figueira and Roy, 2002). The Simos method has found numerous applications and received positive feedback in solving real-life decision problems.

It should be emphasised that other weighting methods are also known; for example, the SWARA (Step-wise Weight Assessment Ratio Analysis) method (Keršuliene et al., 2010; Zolfani et al., 2013) which allows for the inclusion of experts, lawyers or disputed parties' opinions regarding the significance ratio of the attributes in determining rational decisions.

Still other procedures of weighting criteria are described in the literature, e.g., in the following papers: Solymosi and Dombi (1986); Edwards and Barron (1994); Barron and Barret (1996); Roy and Mousseau (1996); Wang and Zionts (2006); De Almeida et al. (2016).

Another frequently used means of determining weights of criteria is the AHP method (Saaty, 1980), which is a relatively common tool for solving multi-criteria decision problems. It uses a hierarchical structure of the decision problem and is implemented in order to:

- indicate the relative significance of criteria (e.g. global preferences, i.e. criteria weights),
- indicate the ratings of decision-making alternatives relative to individual criteria variants (the so-called local preferences).

In the AHP method, preferences on each level of analysis are indicated by means of a pairwise comparison matrix (matrix P) of the factors specified at this level. Pairwise comparison is conducted using Saaty's relative scale of ranks (Saaty, 1980). The ability to indicate the importance of the decision criteria is a substantial benefit of AHP, and the weights determined by its means are used in multi-criteria decision analyses using other methods.

For the determination of the weights, the ANP method has been also proposed. Relevant applications have been described elsewhere (Lin et al., 2010; Kabak et al., 2012). Chiu et al. (2013) proposed to use the DEMATEL-based ANP method for this purpose.

The DEMATEL method was developed in the 1970s with a view to solving complex problems in the identification of cause-effect relationships (Gabus and Fontela, 1972; Fontela and Gabus, 1974). With time, this method has been well adapted for use in multi-criteria decision making. A description of the DEMATEL method can be found elsewhere (e.g. Dytczak, 2010; Michnik, 2013; Tzeng and Huang, 2011).

Some authors discuss the use of DEMATEL to determine the significance of the criteria (e.g. Shieh et al., 2010; Wu and Tsai, 2011; Hsu et al., 2013). Hsu et al. (2013) described using the DEMATEL approach to recognize influential criteria of carbon management in the green supply chain for improving the overall performance of suppliers. Shieh et al. (2010) applied DEMATEL to hospital management by evaluating the importance of criteria and constructing causal relationships among the criteria. Wu and Tsai (2011) discussed the application of DEMATEL to evaluate the importance of criteria in the auto spare parts industry.

In the DEMATEL method, similarly to the AHP/ANP method, structural relationships occur between the analyzed elements. It is a premise for the use of DEMATEL in the weighting of criteria. Some authors have discussed the use of DEMATEL in the weighting process (e.g. Dalalah et al., 2011; Baykasoglu et al., 2013; Patil and Kant, 2014). In some cases, their approach may lead to incorrect results (this will be explained in section 3). Here we propose a new approach to the calculation of criteria weights using DEMATEL and which is different from the procedures used in the papers above. In the next section, a general description of DEMATEL is presented. The subsequent sections deal with the use of DEMATEL in multi-criteria decision analysis, as well as in calculating criteria weights. Furthermore, using numerical examples, a comparison of weights resulting from DEMATEL and AHP has been conducted.

2 Description of the DEMATEL method

As mentioned earlier, the DEMATEL method was elaborated as a procedure for solving problems of identifying cause-and-effect relationships. For modelling problems, DEMATEL uses a direct-influence graph, which expresses the mutual influence of the analysed objects in terms of cause-and-effect relationships (Gabus and Fontela, 1972; Dytczak, 2010; Michnik, 2013; Tzeng and Huang, 2011). Each node of the graph represents an analysed object, whereas an arc between two nodes indicates the direction and intensity of influence relations (Figure 1). The intensity of the influence is defined by values assigned to the given arc. To express the influence of the i-th object on the j-th object, an N-degree scale is used, where: 0 – no influence, 1 – medium influence, …, N – maximum influence. Gabus and Fontela (1972) adopted a 4-degree scale. Currently, the most frequently used are: the original 4-degree scale and a 3-degree scale, but other scales, e.g., a 5-degree scale or even an 8-degree scale are also encountered.

Using the direct influence graph, the direct-influence matrix is created, which is a square matrix whose size is equal to the number of the objects. Its rows correspond to the objects appearing in the comparison as first. The elements on the main diagonal are zeros, while elements $b_{ij}(i \neq j)$ different to zero reflect the impact of the *i*-th object on the *j*-th object:

$$\mathbf{B} = \begin{bmatrix} 0 & b_{1,2} & \dots & b_{1,n} \\ b_{2,1} & 0 & \dots & b_{2,n} \\ \dots & \dots & \dots & \dots \\ b_{n,1} & b_{n,2} & \dots & 0 \end{bmatrix}. \tag{1}$$

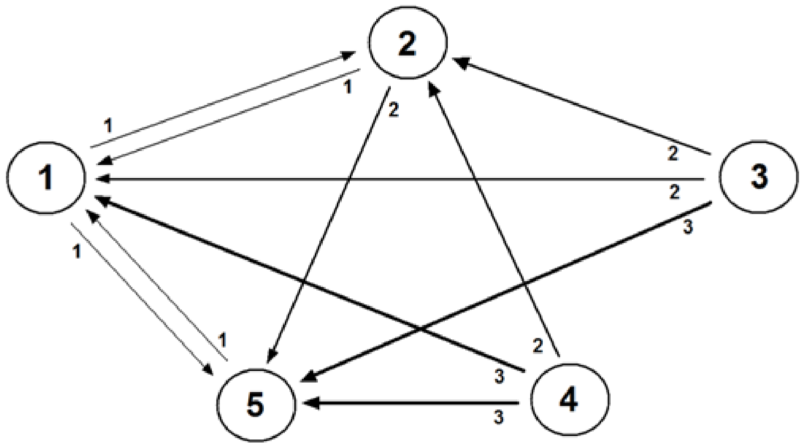


Figure 1. Direct-influence graph in the DEMATEL method

It should be noted that both AHP and DEMATEL are based on the accurate initial matrix, which reflects the relationships between the analysed elements. However, in AHP, the starting point is the pairwise comparison of all elements from each individual level of the structure. As a result, the initial pairwise comparison matrix in the original multiplicative version of AHP does not contain zeros. By contrast, the initial direct influence matrix in DEMATEL does contain zeros. Apart from the main diagonal, zeros can occur also outside the diagonal, if the corresponding objects do not exert sufficient influence on the others.

Matrix (1) is normalised, by dividing each element by the maximum value of the row sum:

$$\hat{\mathbf{B}} = \frac{1}{\max\limits_{i}\left(\sum\limits_{j=1}^{n} b_{i,j}\right)}\mathbf{B} \tag{2}$$

or by the maximum value of the column sum:

$$\hat{\mathbf{B}} = \frac{1}{\max\limits_{j}\left(\sum\limits_{i=1}^{n} b_{i,j}\right)}\mathbf{B}\,. \tag{3}$$

The normalisation of matrix **B** can be performed using the greater of these two sums, which is one of the possible ways of normalisation. The direct influence matrix may also require a more complex normalisation to obtain convergent powers of the matrix in Eq. (5). Here, we use the normalisation recommended by the authors of the DEMATEL method.

From the normalised direct-influence matrix $\hat{\mathbf{B}}$ we calculate the total-influence matrix (**T**), which covers direct and indirect influences $\Delta\hat{\mathbf{B}}$:

$$\mathbf{T} = \hat{\mathbf{B}} + \Delta\hat{\mathbf{B}}\,. \tag{4}$$

The total-influence matrix is described by the equation:

$$\mathbf{T} = \hat{\mathbf{B}} + \hat{\mathbf{B}}^2 + \hat{\mathbf{B}}^3 + \ldots = \hat{\mathbf{B}}\left(\mathbf{I} - \hat{\mathbf{B}}\right)^{-1}, \tag{5}$$

where **I** is the $n \times n$ unit matrix.

Matrix **T** allows to express a relation between the considered objects, covering both direct and indirect influences. For this purpose, appropriate indicators are used, defined as importance indicator ($\boldsymbol{t}^{+}$) and relation indicator ($\boldsymbol{t}^{-}$). They are determined using sums and differences of the row and column sums of matrix **T** corresponding to the *i*-th object:

$$t_i^{+} = \sum_{j=1}^{n} t_{i,j} + \sum_{j=1}^{n} t_{j,i}\,, \tag{6}$$

$$t_i^{-} = \sum_{j=1}^{n} t_{i,j} - \sum_{j=1}^{n} t_{j,i}\,. \tag{7}$$

The importance indicator expresses the role of the object in determining the relation structure between the objects, while the relation indicator expresses the general character of the object, understood as the total influence of this object on all the remaining ones. A positive value of the relation indicator confirms that the given object constitutes the cause, whereas a negative value proves the effect character of the object. The absolute value of the indicator defines the intensity of the effect nature of the object.

DEMATEL can be used as a method of multi-criteria decision making, if the analysed objects represent alternative solutions of the decision problem. Suggestions to use DEMATEL in multi-criteria decision making have been proposed and described in numerous papers (e.g. Chen and Tzeng, 2011; Chen et al., 2010; Lee et al., 2013; Lin and Wu, 2008; Liou, 2007; Shen et al., 2011; Tamura and Akazawa, 2005a, 2005b, 2006; Tzeng et al., 2007; Wu and Lee, 2007; Yang and Tzeng, 2011).

3 The determination of weights using the DEMATEL method

In the papers Baykasoglu et al. (2013) and Dalalah et al. (2011), the DEMATEL method is used also to determine weights of criteria using the following dependencies:

$$\omega_i = \left(\left(t_i^+\right)^2 + \left(t_i^-\right)^2\right)^{1/2}. \tag{8}$$

The values ω_i can be normalised as follows:

$$W_i = \frac{\omega_i}{\sum_{i=1}^{n} \omega_i}, \tag{9}$$

where W_i are the final criteria weights to be used in the decision-making process. This is not an accurate approach, since by using the above equations the same weight is assigned to any two criteria i and j, $i \neq j$, for which $t_i^+ = t_j^+$ but $t_i^- = \left|t_j^-\right|$ (whereby $t_j^- < 0$).

Here, a different approach is proposed for determining the criteria weights using DEMATEL, which does not have this drawback. We assume that the indicators t_i^+ and t_i^- are determined from Eqs. (6) and (7) using the total-influence matrix that results from the direct-influence graph, reflecting the relative importance of the criteria. Since the role and level of the objects' influence are proportional to the value of importance and relation indicators, it is suggested, as one of the possibilities, that the weights be determined as proportional to the average value ($t^{average}$) of the appropriate pair of indicators t_i^+ and t_i^- (Kobryń, 2014). From Eqs. (6) and (7), we obtain:

$$t_i^{average} = \frac{1}{2}\left(t_i^+ + t_i^-\right) = \sum_{j=1}^{n} t_{i,j}. \tag{10}$$

To calculate the normalised weights ($\sum w_i = 1$), the following equation can be used:

$$w_i = \frac{t_i^{average}}{\sum_{i=1}^{n} t_i^{average}} . \tag{11}$$

However, it should be noted that if any criterion is totally dominated by other criteria, the corresponding ratings of this criterion resulting from the direct-influence graph are equal to 0. This creates substantial difficulties, since as a result, the corresponding row in the direct-influence matrix consists of zeros only. Therefore, as seen from Eq. (5), the corresponding row of the total-influence matrix **T** will also consist of zeros. From Eqs. (6), (7) and (10), we see that in this case $t_i^{average} = 0$, and therefore – in accordance with Eq. (11) – we have $w_i = 0$. But this would mean that the given criteria exert practically no influence on results of the analysis.

It may be worth mentioning here that a very popular weighting procedure is the AHP method. However, it should be noted that when a given criterion is dominated by another criterion, the calculation procedure of AHP leads to the assignment of a positive and relatively small weight to this criterion (see numerical examples in the next section).

It is significant that all the criteria should have an adequate influence on the final result of the analysis. Criteria whose weights are zeros cannot occur in the set of criteria. Therefore, when comparing criteria and determining their weights using the DEMATEL method, it is necessary to correct the weight values calculated from Eq. (11).

We propose to increase the weights using the same value δ:

$$w_i^{corrected} = w_i + \delta \tag{12}$$

and then to re-normalise them using the following equation:

$$w_i^{normalized} = \frac{w_i^{corrected}}{\sum_{i=1}^{n} w_i^{corrected}} . \tag{13}$$

The key issue is to determine the correction value δ. Obviously, the influence of the correction on the final weights should be examined.

It seems that the final decision should belong to the decision-maker. Since the main goal is to correct the weight whose initial value is zero, the correction value δ should be as small as possible. The present author suggests setting $\delta \leq \Delta$, where Δ is the smallest non-zero weight of the remaining criteria:

$$\Delta = \min_i w_i \quad \text{if} \quad w_i > 0 . \tag{14}$$

4 Numerical examples

This section presents numerical examples that illustrate the proposed procedure. In addition, the weights calculated using the DEMATEL method were compared with the weights calculated using the very popular AHP method.

Example 1 – determining weights of criteria using the DEMATEL method
Assume that the objects $i = 1, 2, \ldots, 5$ in Figure 1 correspond to the given decision criteria, and that their weights are determined using the suggested procedure. The direct-influence graph from Figure 1 corresponds to the following direct-influence matrix:

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 2 \\ 2 & 2 & 0 & 0 & 3 \\ 3 & 2 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{15}$$

Therefore, the following total-influence matrix **T** is obtained from Eq. (5):

$$\mathbf{T} = \begin{bmatrix} 0.036437 & 0.129555 & 0 & 0 & 0.161943 \\ 0.161943 & 0.020243 & 0 & 0 & 0.275304 \\ 0.348178 & 0.293522 & 0 & 0 & 0.491903 \\ 0.477733 & 0.309717 & 0 & 0 & 0.512146 \\ 0.129555 & 0.016194 & 0 & 0 & 0.020243 \end{bmatrix}. \tag{16}$$

According to Eqs. (6), (7), (10) and (11), the relevant weights are obtained from matrix **T** (Table 1).

Table 1: Determination of criteria weights using DEMATEL

Criterion	Importance (t_i^+)	Relation (t_i^-)	Weight (w_i)
1	1.4818	-0.8259	0.097
2	1.2267	-0.3117	0.135
3	1.1336	1.1336	0.335
4	1.2996	1.2996	0.384
5	1.6275	-1.2955	0.049
		Sum of weights =	1.000

Example 2 – comparison of weights determined using the AHP and DEMATEL methods

Case 2.1: In determining the interdependencies between the criteria, it is important that the relevant ratings be coherent. The coherence of the ratings means that the following conditions are satisfied: $\forall_{i,j,k=1,2,\ldots,m} \quad p_{i,j}\, p_{j,k} = p_{i,k}$, where $p_{i,j}$,

$p_{j,k}$ and $p_{i,k}$ are elements of the pairwise comparison matrix. In practice, those conditions are rarely satisfied by the complete given set of criteria. As it is known, an integral component of the method is the ratings coherence algorithm resulting from the pairwise comparison matrix. There is no such possibility, however, in the DEMATEL method. For that reason, to compare the weights determined using the AHP and DEMATEL methods, we will rely on the ratings assumed for the calculation purposes in the AHP method.

In the case of DEMATEL, the weights are calculated as proposed here, i.e. using Eqs. (10) through (14). Additionally, for comparison, the weights will be calculated using Eqs. (8) and (9).

Assume the following pairwise comparison AHP matrix:

$$\mathbf{P} = \begin{bmatrix} 1 & 9 & 9 & 5 & 3 \\ 1/9 & 1 & 3 & 1/5 & 1/9 \\ 1/9 & 1/3 & 1 & 1/5 & 1/9 \\ 1/5 & 5 & 5 & 1 & 1/3 \\ 1/3 & 9 & 9 & 3 & 1 \end{bmatrix}. \tag{17}$$

The following vector of weights is calculated using AHP:

$$\mathbf{w}^T = [0.505 \quad 0.047 \quad 0.030 \quad 0.132 \quad 0.286]. \tag{18}$$

Following consistency check results are obtained for calculations using AHP: $\lambda_{max} = 5.3691$, $CR = 0.083$. The use of the ratings from matrix **P** in DEMATEL requires the agreement of measurement scales used in both methods. Assuming $N = 8$ in DEMATEL, we can use the scales from Table 2 (Dytczak, 2010).

Table 2: Agreement of measurement scales used in AHP and DEMATEL

Scale level in AHP	**Scale level in DEMATEL (*N* = 8)**
1	0
2	1
3	2
4	3
5	4
6	5
7	6
8	7
9	8

Source: Dytczak (2010).

Additionally, note a lack of feedback for criteria in DEMATEL. For comparison, the following reciprocity rule applies in AHP:

$$p_{j,i} = 1/p_{i,j} \tag{19}$$

If feedback does not occur, as for example between criteria 1 and 2 in Figure 1, the corresponding elements of the direct-influence matrix **B** are zeros. As a result, it can be assumed that matrix **P** in AHP corresponds to the following matrix **B** in DEMATEL:

$$\mathbf{B} = \begin{bmatrix} 0 & 8 & 8 & 4 & 2 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 4 & 0 & 0 \\ 0 & 8 & 8 & 2 & 0 \end{bmatrix}. \qquad (20)$$

From matrix **B**, the following total-influence matrix **T** is obtained:

$$\mathbf{T} = \begin{bmatrix} 0 & 0.431255 & 0.470460 & 0.190083 & 0.090909 \\ 0 & 0 & 0.090909 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.181818 & 0.198347 & 0 & 0 \\ 0 & 0.380165 & 0.414726 & 0.090909 & 0 \end{bmatrix}. \qquad (21)$$

Since criterion 3 is dominated by the remaining criteria (the third row of **B** is filled with zeros), the third row of **T** is also filled with zeros. Therefore, the weight of the third criterion determined from Eq. (10) is $w_3 = 0$. For that reason, a correction of all weights is necessary, as shown in Table 3 (since Δ is small, we set $\delta = \Delta = 0.0358$).

We can see that the weights determined using the two methods usually exhibit high compatibility; the greatest difference of their values occurs for criterion 3 and is equal to $\Delta w = 0.08$.

Table 3: Determining weights of criteria using DEMATEL with correction (case 2.1)

Criterion	**Importance** (t_i^+)	**Relation** (t_i^-)	**Weight** (w_i)	**Corrected weight** $(w_i^{corrected})$	**Normalized weight** $(w_i^{normalizaed})$
1	1.1827	1.1827	0.4657	0.5015	0.425
2	1.0841	-0.9023	0.0358	0.0716	0.061
3	1.1744	-1.1744	0.0000	0.0358	0.031
4	0.6612	0.0992	0.1497	0.1855	0.157
5	0.9767	0.7949	0.3488	0.3846	0.326
		Sum =	1.0000	1.1790	1.000

Case 2.2: Analogous calculations will now be conducted for a different set of initial data. We assume now that the following pairwise comparison matrix is used in the AHP method:

$$\mathbf{P} = \begin{bmatrix} 1 & 1/9 & 1/9 & 1/7 & 1/5 \\ 9 & 1 & 2 & 5 & 5 \\ 9 & 1/2 & 1 & 2 & 3 \\ 7 & 1/5 & 1/2 & 1 & 1/2 \\ 5 & 1/5 & 1/3 & 2 & 1 \end{bmatrix}. \tag{22}$$

From matrix **P**, the following weight vector is obtained:

$$\mathbf{w}^T = \begin{bmatrix} 0.029 & 0.468 & 0.260 & 0.114 & 0.129 \end{bmatrix}. \tag{23}$$

Following consistency check results are obtained for calculations using the AHP method: $\lambda_{max} = 5.2743$, $CR = 0.062$.

Matrix **P** in (22) corresponds in the DEMATEL method to the following matrix **B**:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 8 & 0 & 1 & 4 & 4 \\ 8 & 0 & 0 & 1 & 2 \\ 6 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{24}$$

We obtain the following total-influence matrix **T**:

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.644556 & 0 & 0.058824 & 0.253002 & 0.242215 \\ 0.521474 & 0 & 0 & 0.065744 & 0.117647 \\ 0.352941 & 0 & 0 & 0 & 0 \\ 0.256055 & 0 & 0 & 0.058824 & 0 \end{bmatrix}. \tag{25}$$

In this case, criterion 1 is dominated by the remaining criteria (the first rows of **B** and **T** consist of zeros). Consequently, the weight of the first criterion (w_1) determined from Eq. (9) is equal to 0 (Table 4). It is therefore necessary to correct all weights. Since $\Delta > 0.1$, we set $\delta = 0.5\ \Delta = 0.0612$.

Table 4: Determining weights of criteria using DEMATEL with correction (case 2.2)

Criterion	**Importance** (t_i^+)	**Relation** (t_i^-)	**Weight** (w_i)	**Corrected weight** ($w_i^{corrected}$)	**Normalized weight** ($w_i^{normalizaed}$)
1	1.7750	-1.7750	0.0000	0.0612	0.047
2	1.1986	1.1986	0.4661	0.5274	0.404
3	0.7637	0.6460	0.2741	0.3354	0.257
4	0.7305	-0.0246	0.1373	0.1985	0.152
5	0.6747	-0.0450	0.1225	0.1837	0.140
		Sum =	1.0000	1.3062	1.000

Also in this case, the weights obtained by the two methods exhibit high compatibility (Table 5). As can be seen, the greatest difference between weight values occurs for criterion 2 (Δw = 0.064) and is even smaller than in case 2.1.

Table 5: The weights obtained by AHP and DEMATEL

Case	Criterion	Weights by AHP	Weights by DEMATEL	
			approach outlined in the papers Baykasoglu et al., (2013); Dalalah et al. (2011)	proposed approach
2.1	1	0.505	0.251	0.425
	2	0.047	0.211	0.061
	3	0.030	0.249	0.031
	4	0.132	0.100	0.157
	5	0.286	0.189	0.326
2.2	1	0.029	0.380	0.047
	2	0.468	0.256	0.404
	3	0.260	0.151	0.257
	4	0.114	0.111	0.152
	5	0.129	0.102	0.140

Weights obtained using the proposed approach have been compared also to those obtained using the DEMATEL method as presented in other papers (Baykasoglu et al., 2013; Dalalah et al., 2011). Compared to them, the proposed procedure results in a much higher compatibility of weights with the results obtained by the AHP method.

5 Conclusions

Various methods for obtaining criteria weights are known, such as: the entropy method, the Simos method, the AHP or ANP method, the SWARA method, and many others.

Some authors propose to use for this purpose also the DEMATEL method. This is a popular method, which enables an analysis of cause-and-effect relationships. The potential of this method has also been noted in the context of determining weights of criteria (e.g. Baykasoglu et al., 2013; Dalalah et al., 2011; Hsu et al., 2013; Patil and Kant, 2014; Shieh et al., 2010; Wu and Tsai, 2011). Some of the procedures proposed there, however, have certain drawbacks.

In this paper, a new approach for determining weights of criteria using the DEMATEL method has been presented and verified using numerical examples. Moreover, the obtained weights have been compared to those obtained using other methods, namely AHP and DEMATEL, but following an approach presented elsewhere in the literature (e.g. Baykasoglu et al., 2013; Dalalah et al., 2011). The numerical examples presented here show that the weights determined using the proposed approach exhibit high compatibility with weights determined using the commonly used AHP method.

An implementation of the proposed approach to selected multiple criteria problems may be the next stage of our research. It is a useful method which can be applied to various problems.

References

De Almeida et al. (2016), *A New Method for Elicitation of Criteria Weights in Additive Models: Flexible and Interactive Tradeoff*, European Journal of Operational Research, 250, 179-191.

Barron F.H., Barret B.E. (1996), *Decision Quality Using Ranked Attribute Weights*, Management Science, 42, 1515-1523.

Baykasoglu A., Kaplanoglu V., Durmusoglu Z.D.U., Sahin C. (2013), *Integrating Fuzzy DEMATEL and Fuzzy Hierarchical TOPSIS Methods for Truck Selection*, Expert Systems with Applications, 40, 899-907.

Chen C.H., Tzeng G.H. (2011), *Creating the Aspired Intelligent Assessment Systems for Teaching Materials*, Expert Systems with Applications, 38, 12168-12179.

Chen Y.C., Lien H.P., Tzeng G.H (2010), *Measures and Evaluation for Environment Watershed Plans Using a Novel Hybrid MCDM Model*, Expert Systems and Applications, 37, 926-938.

Chiu W.Y., Tzeng G.H., Li H.L. (2013), *A New Hybrid MCDM Model Combining DANP with VIKOR to Improve e-store Business*, Knowledge-Based Systems, 37, 48-61.

Dalalah D., Hajaneh M., Batieha F. (2011), *A Fuzzy Multi-criteria Decision Making Model for Supplier Selection*, Expert Systems with Applications, 38, 8384-8391.

Deng H., Yeh C.H., Willis R.J. (2000), *Inter-company Comparison Using Modified TOPSIS with Objective Weights*, Computers & Operations Research, 27, 963-973.

Dytczak M. (2010), *Wybrane metody rozwiązywania wielokryterialnych problemów decyzyjnych w budownictwie*, Wydawnictwa Politechniki Opolskiej, Opole.

Edwards W., Barron F.H. (1994), *SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement*, Organizational Behavior and Human Decision Processes, 60, 306-325.

Figueira J., Roy B. (2002), *Determining the Weights of Criteria in the ELECTRE type Methods with a Revised Simos' Procedure*, European Journal of Operational Research, 139, 317-326.

Fontela E., Gabus A. (1974), *DEMATEL, Innovative Methods*, Rep. No. 2, Structural Analysis of the World Problematique (Methods), Battelle Institute, Geneva Research Center, Geneva.

Gabus A., Fontela E. (1972), *World Problems an Invitation to Further thought within the Framework of DEMATEL*, Batelle Institute, Geneva Research Center, Geneva.

Hsu C.W., Kuo T.C., Chen S.H., Hu A.H. (2013), *Using DEMATEL to Develop a Carbon Management Model of Supplier Selection in Green Supply Chain Management*, Journal of Cleaner Production, 56, 164-172.

Ignasiak E., red. (2001), *Badania operacyjne*, PWE, Warszawa.

Kabak M., Burmaoglu S., Kazancoglu Y. (2012), *A Fuzzy Hybrid MCDM Approach for Professional Selection*, Expert Systems with Applications, 39, 3516-3525.

Keršuliene V., Zavadskas E.K., Turskis Z. (2010), *Selection of Rational Dispute Resolution Method by Applying New Step-wise Weight Assessment Ratio Analysis (SWARA)*, Journal of Business Economics and Management, 11(2), 243-258.

Kobryń A. (2014), *Wielokryterialne wspomaganie decyzji w gospodarowaniu przestrzenią*, Wydawnictwo DIFIN, Warszawa.

Lee H.S., Tzeng G.H., Yeih W., Wang Y.J., Yang S.C. (2013), *Revised DEMATEL: Resolving the Infeasibility of DEMATEL*, Applied Mathematical Modelling, 37, 6746-6757.

Lin C.J., Wu W.W. (2008), *A Causal Analytical Method for Group Decision-making under Fuzzy Environment*, Expert Systems and Applications, 34, 205-213.

Lin Y.T., Lin C.L., Yu H.C., Tzeng G.H. (2010), *A Novel Hybrid MDCM Approach for Outsourcing Vendor Selection: A Case Study for a Semiconductor Company in Taiwan*, Expert Systems and Applications, 37, 4796-4804.

Liou Y. (2007), *Water Quality Analysis Model Based on DEMATEL and ANP*, 22nd European Conference on Operational Research EURO XII. Book of Abstracts, Praha.

Michnik J. (2013), *Wielokryterialne metody wspomagania decyzji w procesie innowacji*, Prace Naukowe Uniwersytetu Ekonomicznego, Katowice.

Patil S.K., Kant R. (2014), *A Hybrid Approach Based on Fuzzy DEMATEL and FMCDM to Predict Success of Knowledge Management Adoption in Supply Chain*, Applied Soft Computing, 18, 126-135.

Roy R., Mousseau V. (1996), *A Theoretical Framework for Analyzing the Notion of Relative Importance of Criteria*, Journal of Multi-Criteria Decision Analysis, 5, 145-159.

Saaty T.L. (1980), *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill, New York.

Shannon C.E., Weaver W. (1963), *The Mathematical Theory of Communication*, University of Illinois Press, Urbana and Chicago.

Shen Y.C., Lin G.T., Tzeng G.H. (2011), *Combined DEMATEL Techniques with Novel MCDM for the Organic Light Emitting Diode Technology Selection*, Expert Systems with Applications, 38, 1468-1481.

Shieh J.I., Wu H.H., Huang K.H. (2010), *A DEMATEL Method in Identifying Key Success Factors of Hospital Service Quality*, Knowledge-Based Systems, 23, 277-282.

Simos J. (1990), *L'évaluation environnementale*: *Un processus cognitive négocié, Thèse de doctorat*, DGF-EPFL, Lausanne.

Solymosi T., Dombi J. (1986), *A Method for Determining the Weights of Criteria: The Centralized Weights*, European Journal of Operational Research, 26, 35-41.

Tamura H., Akazawa K. (2005a), *Structural Modeling and Systems Analysis of Uneasy Factors for Realizing Safe, Secure and Reliable Society*, Journal of Telecommunications and Information Technology, 3, 64-72.

Tamura H., Akazawa K. (2005b), *Stochastic DEMATEL for Structural Modeling of a Complex Problematique for Realizing Safe, Secure and Reliable Society*, Journal of Telecommunications and Information Technology, 4, 139-146.

Tamura H., Akazawa K. (2006), *Decision Support for Extracting and Dissolving Consumers' Uneasiness over Foods Using Stochastic DEMATEL*, Journal of Telecommunications and Information Technology, 4, 91-95.

Tzeng G.H., Chiang C.H., Li C.W. (2007), *Evaluating Intertwined Effects in e-learning Programs: A Novel Hybrid MCDM Model Based on Factor Analysis and DEMATEL*, Expert Systems with Applications, 32(4), 1028-1044.

Tzeng G.H., Huang J.J. (2011), *Multiple Attribute Decision Making: Methods and Applications*, CRC Press, Taylor & Francis Group.

Wang J., Zionts S. (2006), *Random-weight Generation in Multiple Criteria Decision Models*, MCDM' 2006, Chania, Greece, June 19-23.

Wu W.W., Lee Y.T. (2007), *Developing Global Managers' Competencies Using the Fuzzy DEMATEL Method*, Expert Systems with Applications, 32, 499-507.

Wu H.H, Tsai Y.N. (2011), *A DEMATEL Method to Evaluate the Causal Relations among the Criteria in Auto Spare Parts Industry*, Applied Mathematics and Computation, 218, 2334-2342.

Yang J., Tzeng G.H. (2011), *An Integrated MCDM Technique Combined with DEMATEL for a Novel Cluster-weighted with ANP Method*, Expert Systems and Applications, 38, 1417-1424.

Zolfani S.H., Aghdaie M.H., Derakthi A., Zavadskas E.K., Varzandeh M.H.M. (2013), *Decision Making on Business Issues with Foresight Perspective: An Application of New Hybrid MCDM Model in Shopping Mall Locating*, Expert Systems and Applications, 40, 7111-7121.

M U L T I P L E   C R I T E R I A   D E C I S I O N   M A K I N G
Vol. 12 2017

Maciej Nowak[*]
Sebastian Sitarz[**]
Tadeusz Trzaskalik[***]

INTERACTIVE PROCEDURE FOR MULTIOBJECTIVE DYNAMIC PROGRAMMING WITH THE MIXED ORDERED STRUCTURE

DOI: 10.22367/mcdm.2017.12.12

Abstract

The paper presents a multiobjective dynamic programming problem with the values of the criteria function in ordered structures. The first problem is a model with deterministic values; the second, one with triangular fuzzy numbers; and the third, one with discrete random variables with the *k*-th absolute moment finite. The fourth model is a product of the three models listed above. The aim of the paper is to present an interactive procedure which uses trade-offs and which allows to determine the final solution in the mixed ordered structure. The ordered structures and the proposed procedure are illustrated by numerical examples.

Keywords: multiobjective dynamic programming, interactive procedures, partially ordered criteria space, mixed partially ordered structures.

1 Introduction

Multiobjective, multicriteria decision problems are usually investigated as models of multicriteria dynamic programming, using the vector version of Bellman's principle of optimality (1957), non-dominated evaluations (in the criteria space) and efficient solutions (in the decision space). An example of this approach can

[*] University of Economics in Katowice, Faculty of Informatics and Communication, Department of Operations Research, Katowice, Poland, e-mail: maciej.nowak@ue.katowice.pl.
[**] University of Silesia, Institute of Mathematics, Katowice, Poland, e-mail: ssitarz@math.us.edu.pl.
[***] University of Economics in Katowice, Faculty of Informatics and Communication, Department of Operations Research, Katowice, Poland, e-mail: tadeusz.trzaskalik@ue.katowice.pl.

be found in the papers by Trzaskalik (1991) for the deterministic case, and by Trzaskalik and Hoa (1999) for the stochastic case. Another group of problems consists of scalarization problems which allow to transform the given multicriteria problem into the corresponding single-criterion problem. An overview bibliography on this topic can be found in the papers by Li and Haimes (1989) and by Trzaskalik (1998).

Another method of generalization of single-criterion dynamic programming consists in regarding the evaluations as elements of a partially ordered space. First papers on this subject were written by Mitten (1974) and Sobol (1975), Steinberg and Parks (1979), and Henig (1985). Discrete dynamic programming with evaluations in a partially ordered space was also considered in the papers by Trzaskalik and Sitarz (2002; 2007).

A problem that appears in many decision models is that of the simultaneous occurrence of deterministic, stochastic, and fuzzy values in the set of multidimensional evaluations. Such situations are described by Zaraś (2004). A question arises: Can such mixed evaluations be used in optimal control of a multiobjective decision process according to a homogeneous scheme in ordered structures? This issue was discussed in detail in the paper by Trzaskalik and Sitarz (2004). The authors considered first the situation with a homogeneous one- or multidimensional evaluation space, consisting of real numbers, triangular fuzzy numbers, or first-order stochastic dominances. Further in the paper, examples of combinations of such structures are given, allowing to obtain product structures which are also ordered structures.

The fundamental problem in multicriteria decision making is the selection of the final solution. Commonly used for this purpose are interactive methods, which allow to include the decision maker in the process of obtaining the final decision. Worth mentioning here are selected interactive methods described in the literature. Benayoun et al. (1971) suggested the STEM method, consisting in reducing the set of admissible solutions using the Chebyshev metric. Steuer (1977) described an interactive method based on the determination of weight intervals for the criteria, to obtain a reduced criteria cone. The essence of the Korhonen and Laakso method (1986), on the other hand, is computer visualization of the information on the final solution, obtained in the consecutive iterations. Miettinen and Makela (2000) designed the NIMBUS method, which operates together with the decision maker using the Internet. In this method, the decision maker divides the obtained solutions into five classes describing his/her preferences. In turn, Ozpeynirci et al. (2017) constructed an interactive method based on narrowing the criteria cone by pairwise comparisons of selected admissible solutions.

Interactive methods are used mainly to solve deterministic multicriteria problems. It happens often, however, that the data available to the decision maker do not allow to formulate the problem in such categories. Interactive methods for decision making under risk, when evaluations of alternatives are expressed by probability distributions, have been proposed, for instance, in the papers by Nowak (2006; 2007; 2010). In the last paper, trade-offs are used to determine a new candidate solution. A similar approach is used in the present paper. The paper by Nowak and Trzaskalik (2013), on the other hand, presents an interactive approach to the dynamic decision-making problem under risk.

Trzaskalik and Sitarz (2004) focused on finding maximal solutions in various ordered structures. In the present paper, which is a continuation of the previous paper, we discuss finding the final solution in a mixed model with deterministic, stochastic, and fuzzy criteria. The goal of the present paper is to propose an interactive procedure with trade-offs, which allows to determine the final solution.

Further on, in Section 2, we discuss a multiobjective decision process in an ordered structure, as well as the problem of finding the set of maximal evaluations and the corresponding set of efficient realizations. In Section 3 we present three ordered structures: one with the set of reals as the fundamental set, another one with the set of triangular fuzzy numbers, and the third one with the set of discrete random variables with the *k*-th absolute moment finite. Next, we present a structure which is a product of the three ordered structures listed above. In the illustrative example in that section we find all the maximal values and the corresponding efficient realizations. In Section 4 we describe the proposed interactive procedure which allows to find the final solution. Section 5 is a summary of the paper.

2 Multiobjective decision process in an ordered structure

We consider a finite, discrete dynamic Markov process whose sets of states and decisions at each stage are finite, and whose transfer function is deterministic. A stage realization is defined as a pair consisting of a process state and a feasible decision. A process realization is a sequence of stage realizations such that the state at the beginning of the next stage is a consequence of the stage realizations at the previous stage, described by the transfer function. Each discrete multiobjective decision process can be assigned a graph whose vertices are process states and edges are decisions. Each process realization corresponds to a path joining two vertices.

Example 1
An example of a multiobjective process is shown in Figure 1.

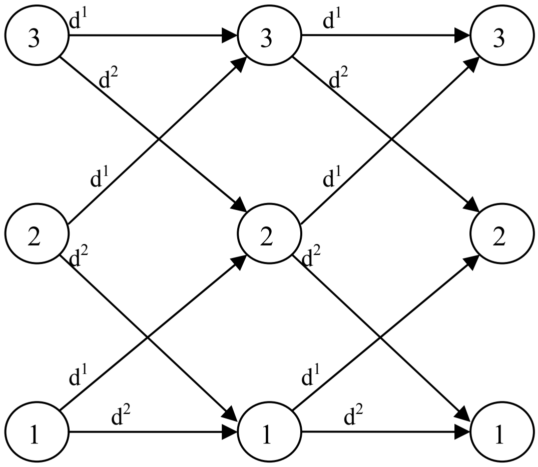


Figure 1. An example of a dynamic process

We consider a two-stage dynamic process with three admissible states at the beginning of each stage: 1, 2, 3. In each state, we can make one of the following two decisions: d^1, d^2. Depending on our current state and the decision made, we proceed to the next state according to the transfer function:

$$\Omega_t(s_t, d_t) = s_{t+1} \tag{1}$$

which is given by the following formulas:

$$\Omega_t(1, d^1) = 2, \Omega_t(1, d^2) = 1$$
$$\Omega_t(2, d^1) = 3, \Omega_t(2, d^2) = 1$$
$$\Omega_t(3, d^1) = 3, \Omega_t(3, d^2) = 2$$

To each stage realization of the process we assign a stage evaluation, while the combined evaluation of the process is an aggregate of stage evaluations. Stage and multistage evaluations are elements of a space W. We assume that a binary operator $\circ$ combining stage evaluations, as well as an ordering relation $\preccurlyeq$, are defined in W. Let $a, b \in$ W. If $a \preccurlyeq b$, we say that element b is not worse than element a.

The structure (W, $\circ$, $\preccurlyeq$) is called an ordered structure, if it satisfies the following condition (further referred to as TS):

$$\forall_{a \in \mathrm{W}}\ a \preccurlyeq a \tag{2}$$
$$\forall_{a,b \in \mathrm{W}}\ a \preccurlyeq b\ \wedge b \preccurlyeq a\ \Longrightarrow a = b \tag{3}$$
$$\forall_{a,b,c \in \mathrm{W}}\ a \circ (b \circ c) = (a \circ b) \circ c \tag{4}$$

$$\forall_{a,b,c \in \mathrm{W}}\ a \preccurlyeq b \Rightarrow a \circ c \preccurlyeq b \circ c \wedge c \circ a \preccurlyeq c \circ b \tag{5}$$

Using relation $\preccurlyeq$ we define the following relation $\prec$:

$$a \prec b \Leftrightarrow a \preccurlyeq b \wedge a \neq b \tag{6}$$

Let $a, b \in \mathrm{W}$. If $a \prec b$, we say that element b is better than element a. Using relation $\prec$ we determine maximal elements of set $\mathrm{A} \subset \mathrm{W}$:

$$\max(\mathrm{A}) = \{a^* \in \mathrm{A}:\ \sim \exists_{a \in \mathrm{A}}\, a^* \prec a\} \tag{7}$$

In what follows, we will use the following notational convention:

$$\max\{a_1, \dots, a_k\} = \max(\{a_1, \dots, a_k\}) \tag{8}$$

We want to find all maximal evaluations of the process and the corresponding realizations, called efficient realizations. For this purpose, we use dynamic programming and Bellman's principle of optimality. A formal description of the procedure can be found in the papers by Trzaskalik and Sitarz (2002; 2007).

3 Examples of ordered structures

3.1 Structure S_1: $(\Re, +, \leq)$

As the first structure, we will consider a structure with the set of real numbers as set W. This is illustrated by a process with the same sets of admissible states and decisions, and the same transfer function as in Example 1.

Example 2

We consider a two-stage process shown in Figure 2. The values on the edges are real numbers expressing stage evaluations of the process.

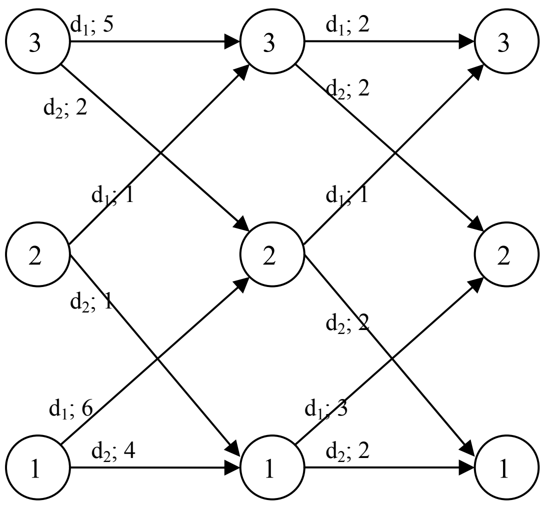


Figure 2. An illustration of the ordered structure S_1

Using dynamic programming (starting with the last stage) we find the maximal values of partial realizations of the process which start at the given state and which proceed until the end of the process, as well as the corresponding decisions. Next, we find the maximal value of the process, which is equal to 8, and the corresponding efficient realization (1, d_1, 2, d_2). Detailed calculations can be found in Trzaskalik and Sitarz (2004).

3.2 Structure S_2: (W_F, $+_F$, $\leq_F$)

Our next ordered structure is the structure (W_F, $+_F$, $\leq_F$), where:

$$W_F = \{(m, \alpha, \beta): m \in \Re, \alpha > 0, \beta > 0\} \quad (9)$$

is a set of triangular fuzzy numbers, where m is the center of the fuzzy number, and α, β are its spreads. The operator $+_F$ combining the values of the criteria function is the sum of triangular fuzzy numbers (m_1, α_1, β_1), (m_2, α_2, β_2) and is defined as follows:

$$(m_1, \alpha_1, \beta_1) +_F (m_2, \alpha_2, \beta_2) = (m_1 + m_2, \alpha_1 + \alpha_2, \beta_2 + \beta_1) \quad (10)$$

The ordering relation $\leq_F$ is defined as follows:

$$(m_1, \alpha_1, \beta_1) \leq_F (m_2, \alpha_2, \beta_2) \Leftrightarrow (m_1 \leq m_2 \wedge m_1 - \alpha_1 \leq m_2 - \alpha_2 \wedge m_1 + \beta_1 \leq m_2 + \beta_2) \quad (11)$$

This is illustrated by a process with the same sets of admissible states and decisions, and the same transfer function, as in Example 1.

Example 3

We consider a two-stage process shown in Figure 3. The values on the edges are triangular fuzzy numbers, expressing stage evaluations of the process.

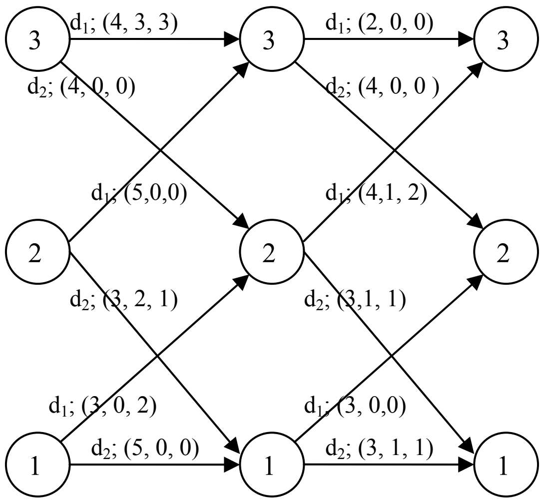


Figure 3. An illustration of the ordered structure S_2.

Using dynamic programming we obtain the following efficient realizations and the corresponding maximal values:

$(1, d_1, 2, d_1) \to (7, 1, 4)$ $(2, d_1, 3, d_2) \to (9, 0, 0)$

$(3, d_1, 3, d_2) \to (8, 3, 3)$ $(3, d_2, 2, d_1) \to (8, 1, 2)$

Detailed calculations can be found in Trzaskalik and Sitarz (2004).

3.3 Structure S_3: $(W_R, +_R, \leq_R)$

In this structure, the set W_R contains discrete random variables with the k-th absolute moment finite. We consider discrete random variables which can admit only a finite number of values from the set {0, 1, 2, ...}. The set of such random variables can be expressed as a set of probability sequences:

$$W_R = \{(p_0, p_1, p_2, ..., p_n): p_n > 0, p_i \geq 0, \sum_{i=0}^{n} p_i = 1\} \tag{12}$$

where p_i is the probability of the number $i \geq 0$.

As the operator $\circ$ we take addition of random variables.

Let $p = (p_0, p_1, p_2, ..., p_n)$, $q = (q_0, q_1, \ldots, q_m)$; then $p \circ q$ is defined as follows:

$$p +_R q = (r_0, r_1, ..., r_{n+m}) \tag{13}$$

where $r_i = \sum_{k,l:\ k+l=i} p_k q_l$.

The ordering relation is determined by the first-order stochastic dominance which can be characterized as follows. Let $p = (p_0, p_1, ..., p_n)$, $q = (q_0, q_1, \ldots, q_m)$.

Then:

$$p \leq_R q \Leftrightarrow (\forall_{i=1,\ldots,\max\{n,m\}} P_i \geq Q_i) \tag{14}$$

where $P_i = \sum_{k=0}^{i} p_k$, $Q_i = \sum_{k=0}^{i} q_k$.

This is illustrated by a process with the same sets of admissible states and decisions, and the same transfer function, as in Example 1.

Example 4

We consider a two-stage process shown in Figure 4. The values on the edges are discrete random variables which can admit only a finite number of values from the set {0, 1, 2, ...}.

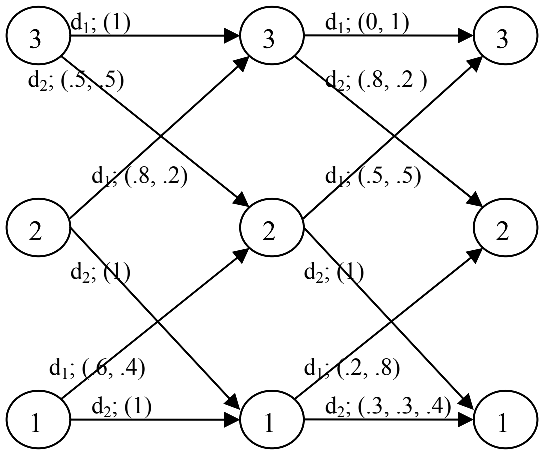


Figure 4. An illustration of the ordered structure S_3

Analogously as in the previous examples, we obtain the following efficient realizations and the corresponding maximal values:

$$(1, d_2, 1, d_2) \to (.2,.8) \qquad (2, d_2, 1, d_2) \to (.3, .3, .4)$$
$$(2, d_1, 3, d_2) \to (0, .8, .2) \qquad (3, d_1, 3, d_2) \to (.25, .5, .25)$$

Detailed calculations can be found in Trzaskalik and Sitarz (2004).

3.4 Structure S_4: $[(\Re, W_F, W_R), (+, +_F, +_R), (\leq, \leq_F, \leq_R)]$

Structure S_4 is a product of the three structures: with real numbers, with triangular fuzzy numbers, and with random variables with stochastic dominances. Let $[a_1, (m_1, \alpha_1, \beta_1), p_1], [a_2, (m_2, \alpha_2, \beta_2), p_2] \in [(\Re, W_F, W_R)$. Then:

$$\begin{aligned}&[a_1, (m_1, \alpha_1, \beta_1), p_1]\ (+, +_F, +_R)\ [a_2, (m_2, \alpha_2, \beta_2), p_2] = \\ &= [a_1 + a_2, (m_1, \alpha_1, \beta_1) +_F (m_2, \alpha_2, \beta_2), p_1 +_R p_2]\end{aligned} \tag{15}$$

$$\begin{aligned}&[a_1, (m_1, \alpha_1, \beta_1), p_1]\ (\leq, \leq_F, \leq_R)\ [a_2, (m_2, \alpha_2, \beta_2), p_2] \Leftrightarrow \\ &\Leftrightarrow a_1 \leq a_2 \wedge (m_1, \alpha_1, \beta_1) \leq_F (m_2, \alpha_2, \beta_2) \wedge p_1 \leq_R p_2\end{aligned} \tag{16}$$

This case is also illustrated by a process with the same sets of admissible states and decisions, and the same transfer function, as in Example 1.

Example 5
Consider the dynamic process shown in Figure 5. The values on the edges are real numbers, triangular fuzzy numbers, and random variables. Using dynamic programming we obtain efficient realizations and the corresponding maximal values, listed together in Table 1:

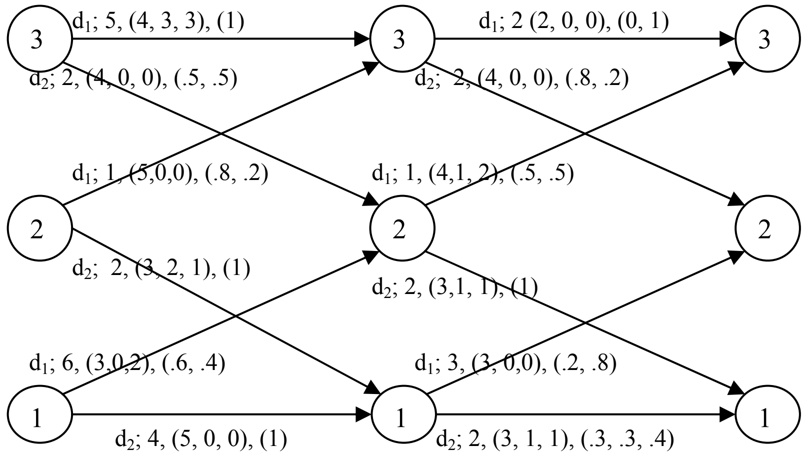


Figure 5. An illustration of the ordered structure S_4

Table 1: Efficient realizations and maximal values for structure S_4

No.	Realization	(ℜ, fuzzy number, random variable)
1	(1, d_2, 1, d_2)	6, (8, 1, 1), (.3, .3, .4)
2	(1, d_2, 1, d_1)	7, (8, 0, 0), (.2, .8)
3	(1, d_1, 2, d_2)	8, (6, 1, 3), (.6, .4)
4	(1, d_1, 2, d_1)	7, (7, 1, 4), (.3, .5, .2)
5	(2, d_2, 1, d_2)	3, (6, 3, 2), (.3, .3, .4)
6	(2, d_1, 3, d_2)	5, (9, 0, 0), (.64, .34, .02)
7	(2, d_1, 3, d_1)	5, (7, 0, 0), (0, .8, .2)
8	(3, d_2, 2, d_1)	3, (8, 1, 2), (.25, .5, .25)
9	(3, d_1, 3, d_2)	8, (8, 3, 3), (.8, .2)
10	(3, d_1, 3, d_1)	7, (6, 3, 3), (0, 1)

4 Interactive procedure

The set of efficient realizations is usually so large that choosing one of them as the final solution can be difficult. Therefore, we suggest applying an interactive procedure which enables the decision maker to identify the solution best suited to his/her expectations.

To facilitate the analysis of the solutions presented to the decision maker in the consecutive iterations of the procedure, we will scalarize the evaluations of efficient realizations. That is, we will transform them so that each evaluation will be given as a real number.

The set of real numbers $\Re$ already consists of scalars, therefore we take the identity as the scalarization:

$$a \to a$$

The set of fuzzy numbers W_F consists of triples of real numbers: center and two spreads (left and right). As the scalarization we will take the map assigning to each fuzzy number its center:

$$(m, \alpha, \beta) \to m$$

For a random variable from the set W_R, as the scalarization we will take the expected value:

$$p \to E(p)$$

Let $D = \{d_1, d_2, \ldots, d_m\}$ be the set of efficient realizations of the analyzed process and $F = \{f_1, f_2, \ldots, f_n\}$ the set of criteria used for evaluation. The evaluation of each realization with respect to the criteria can be given as a real number, a discrete random variable with the *k*-th absolute moment finite, or as a triangular fuzzy number. We denote by $f_j(d_i)$ the scalarized evaluation of realization d_i with respect to criterion f_j.

In the procedure described below, we will also use standardized evaluations of the realizations with respect to criteria $g_j(d_i)$, which will be determined from the following formula:

$$g_j(d_i) = \frac{f_j(d_i) - \min_{i \in \overline{1,m}}\{f_j(d_i)\}}{max_{i \in \overline{1,m}}\{f_j(d_i)\} - min_{i \in \overline{1,m}}\{f_j(d_i)\}} \tag{17}$$

Let $D^{(l)}$ be the set of realizations considered in iteration l. In each iteration of the interactive procedure the DM is shown a certain candidate realization $d^{(l)}$ and a potency matrix $M^{(l)}$ with two rows: the first one groups the largest values of the criteria used for the realizations from set $D^{(l)}$, and the second one, the smallest values:

$$M^{(l)} = \begin{bmatrix} \overline{f}_1^{(l)} & \cdots & \overline{f}_n^{(l)} \\ \underline{f}_1^{(l)} & \cdots & \underline{f}_n^{(l)} \end{bmatrix} \tag{18}$$

where:

$$\overline{f}_j^{(l)} = \max_{d_i \in D^{(l)}} f_j(d_i), j \in \overline{1,n} \tag{19}$$

$$\underline{f}_j^{(l)} = \min_{d_i \in D^{(l)}} f_j(d_i), j \in \overline{1,n} \tag{20}$$

The proposed interactive procedure consists of the following steps:

Preliminary stage:

1. For each efficient realization, calculate the scalar values with respect to all criteria $f_j(d_i)$.
2. Using formula (17), calculate the standardized values of the evaluations of efficient realizations with respect to criteria $g_j(d_i)$.
3. Determine the first candidate realization $d^{(1)}$ using the min-max criterion:
   a) For each realization, determine the minimum of the standardized evaluations with respect to each criterion:
   $$g^{\min}(d_i) = \min_{j\in\overline{1,n}}\{g_j(d_i)\} \quad (21)$$
   b) As the first candidate realization $d^{(1)}$ take that d_i for which the value $g^{\min}(d_i)$ is maximal.
4. Set $l = 1$ and $\mathrm{D}^{(1)} = \mathrm{D}$ and proceed to the first iteration.

Iteration *l*

1) Determine the potency matrix $\mathrm{M}^{(l)}$.
2) Present the values of the criteria obtained for realization $d^{(l)}$ and potency matrix $\mathrm{M}^{(l)}$ to the DM. If the DM is satisfied with the proposed realization, end the procedure.
3) Ask the DM to assign each criterion to one of the following three sets:
   F_1 – the set of criteria whose values should be improved as compared with the value obtained for realization $d^{(l)}$,
   F_2 – the set of criteria whose values should not be made worse as compared with the value obtained for realization $d^{(l)}$,
   F_3 – the set of criteria whose values can be made worse as compared with the value obtained for realization $d^{(l)}$.
4) Determine the set $\mathrm{D}^{(l+1)}$ consisting of all the realizations from the set $\mathrm{D}^{(l)}$ which satisfy the following conditions:
   $$\forall_{f_j\in\mathrm{F}_1}\ f_j(d_i) > f_j\left(d^{(l)}\right) \quad (22)$$
   $$\forall_{f_j\in\mathrm{F}_2}\ f_j(d_i) \geq f_j\left(d^{(l)}\right) \quad (23)$$
5) If $\mathrm{D}^{(l+1)} = \emptyset$, inform the decision maker that no realization exists for which the values of the criteria from F_1 are higher than for realization $d^{(l)}$, and the values of the criteria from F_2 are not lower than those for realization $d^{(l)}$. Return to step (2).
6) If $\mathrm{D}^{(l+1)}$ consists of one realization only, take this realization as the next proposed realization $d^{(l+1)}$. Proceed to step (10).

7) For each realization $d_i \in \mathrm{D}^{(l+1)}$ and for each criteria pair (f_j, f_k), such that $f_j \in \mathrm{F}_1, f_k \in \mathrm{F}_3$ and $f_k(d_i) < f_k(d^{(l)})$, calculate the value of the trade-off $t_{jk}(d_i)$ from the formula:

$$t_{jk}(d_i) = \frac{g_j(d_i) - g_j(d^{(l)})}{g_k(d^{(l)}) - g_k(d_i)} \tag{24}$$

8) For each criteria pair (f_j, f_k) such that $f_j \in \mathrm{F}_1, f_k \in \mathrm{F}_3$, check if there exists at least one realization $d_i \in \mathrm{D}^{(l+1)}$, for which the value of $t_{jk}(d_i)$ has been calculated in step 5. If so, then for each realization $d_m \in \mathrm{D}^{(l+1)}$ such that $f_k(d_m) \geq f_k(d^{(l)})$, take as the trade-off $t_{jk}(d_i)$ twice the maximal value of the trade-offs calculated for the pair (f_j, f_k) in step 5. If for every realization we have $d_m \in \mathrm{D}^{(l+1)}$, take $t_{jk}(d_m) = 1$.
9) For each realization $d_i \in \mathrm{D}^{(l+1)}$, calculate the average of the trade-offs calculated in steps 5 and 6 for each criteria pair (f_j, f_k), such that $f_j \in \mathrm{F}_1, f_k \in \mathrm{F}_3$. As the next realization $d^{(l+1)}$ to be proposed to the decision maker take the one for which this average is highest.
10) Set $l = l + 1$ and proceed to the next iteration.

The first candidate realization is determined using the min-max criterion. In each iteration, the DM is presented with evaluations of the proposed realization and with the potency matrix, which consists of maximal and minimal criteria values obtained for the currently considered realizations. The DM can either accept the proposed realizations as the solution of the problem, or else determine the direction of improvement, by indicating:

a) which criteria should achieve a value higher than the one obtained for the candidate realization,
b) which criteria should retain the value obtained for the candidate realization,
c) which criteria can have a lower value than the one obtained for the candidate realization.

Of course, since we operate within the set of efficient realizations, the decision maker must indicate at least one criterion whose value can be lowered.

The procedure should continue until the decision maker is satisfied with the proposed realization (step 2). During the dialog it can turn out, however, that the consecutive proposals do not satisfy the decision maker's expectations. He/she can then either end the procedure or else consider once again the realizations proposed earlier and decide to select one of them.

5 An illustration of the interactive procedure

Consider the problem from Example 5. In the preliminary stage, we calculate the values of the scalarized evaluations of efficient realizations (Table 2) and standardized values (Table 3).

Table 2: Values of scalarized evaluations for efficient realizations

d_i	$f_1(d_i)$	$f_3(d_i)$	$f_3(d_i)$
d_1 (1, 2, 1, 2)	6	8	1.1
d_2 (1, 2, 1, 1)	7	8	0.8
d_3 (1, 1, 2, 2)	8	6	0.4
d_4 (1, 1, 2, 1)	7	7	0.9
d_5 (2, 2, 1, 2)	3	6	1.1
d_6 (2, 1, 3, 2)	5	9	0.4
d_7 (2, 1, 3, 1)	5	7	1.2
d_8 (3, 2, 2, 1)	3	8	1.0
d_9 (3, 1, 3, 2)	8	8	0.2
d_{10} (3, 1, 3, 1)	7	6	1.0
Min	3	6	0.2
Max	8	9	1.2

Table 3: Values of standardized evaluations for efficient realizations

d_i	$g_1(d_i)$	$g_3(d_i)$	$g_3(d_i)$	Min
d_1 (1, 2, 1, 2)	0.60	0.67	0.90	0.60
d_2 (1, 2, 1, 1)	0.80	0.67	0.60	0.60
d_3 (1, 1, 2, 2)	1.00	0.00	0.20	0.00
d_4 (1, 1, 2, 1)	0.80	0.33	0.70	0.33
d_5 (2, 2, 1, 2)	0.00	0.00	0.90	0.00
d_6 (2, 1, 3, 2)	0.40	1.00	0.20	0.20
d_7 (2, 1, 3, 1)	0.40	0.33	1.00	0.33
d_8 (3, 2, 2, 1)	0.00	0.67	0.80	0.00
d_9 (3, 1, 3, 2)	1.00	0.67	0.00	0.00
d_{10} (3, 1, 3, 1)	0.80	0.00	0.80	0.00

We include all the efficient realizations in set $D^{(1)}$:

$$D^{(1)} = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}\}$$

As the first candidate realization, we take d_1:

$$d^{(1)} = d_1$$

The following calculations are performed in the consecutive iterations:

Iteration 1

1) Determine the potency matrix $M^{(1)}$.
2) Present the criteria values for realization $d^{(1)}$ and the potency matrix $M^{(1)}$ (Table 4) to the DM.

Table 4: Candidate realization and potency matrix from iteration 1

Criterion	f_1	f_2	f_3
$d^{(1)}$	6	8	1.1
Minimal value	3	6	0.2
Maximal value	8	9	1.2

The DM is not satisfied with realization $d^{(1)}$.

3) The DM decides that the value of criterion f_1 should be higher than 6, while the values of the remaining criteria can be lowered as compared with the ones obtained for realization $d^{(1)}$:

$$\mathrm{F}_1 = \{ f_1 \}, \mathrm{F}_2 = \emptyset, \mathrm{F}_3 = \{ f_2, f_3\}$$

4) Determine the set of variants satisfying the condition formulated by the DM in step 3:

$$\mathrm{D}^{(2)} = \{d_2, d_3, d_4, d_9, d_{10}\}$$

5) Since $\mathrm{D}^{(2)} \neq \emptyset$, proceed to the next step.
6) Since $\mathrm{D}^{(2)}$ contains more than one realization, proceed to the next step.
7) Calculate trade-offs $t_{12}(d_i)$ and $t_{13}(d_i)$ for $d_i \in \mathrm{D}^{(2)}$. When calculating the first value, omit the realizations for which $f_2(d_i) \geq f_2(d^{(1)})$, that is, d_2 and d_9. For the remaining realizations from $\mathrm{D}^{(2)}$ the trade-offs are:

$$t_{12}(d_3) = 0.60, t_{12}(d_4) = 0.60, t_{12}(d_{10}) = 0.30$$

When calculating the trade-offs, we note that $f_3(d_i) \geq f_3(d^{(1)})$ does not hold for any $d_i \in \mathrm{D}^{(2)}$. The trade-offs are:

$$t_{13}(d_2) = 0.67, t_{13}(d_3) = 0.57, t_{13}(d_4) = 1.00, t_{13}(d_9) =$$
$$= 0.44, t_{12}(d_{10}) = 2.00$$

8) For d_2 and d_9, for which trade-offs $t_{12}(d_i)$ have not been calculated in step 7, take:

$$t_{12}(d_2) = t_{12}(d_9) = 2 \cdot \max\{t_{12}(d_3), t_{12}(d_4), t_{12}(d_{10})\} = 1.20$$

9) Calculate the average values of the trade-offs (Table 5).

Table 5: Values of the trade-offs calculated in iteration 1

d_i	$t_{12}(d_i)$	$t_{13}(d_i)$	Average
d_2	1.20	0.67	0.93
d_3	0.60	0.57	0.59
d_4	0.60	1.00	0.80
d_9	1.20	0.44	0.82
d_{10}	0.30	1.00	1.15

The next realization $d^{(2)}$ proposed to the DM is d_{10}.

10) Set $l = 2$ and proceed to the next iteration.

Iteration 2

1) Determine the potency matrix $\mathrm{M}^{(2)}$.
2) Present the criteria values obtained for realization $d^{(2)}$ and the potency matrix $\mathrm{M}^{(2)}$ (Table 6) to the DM.

Table 6: Candidate realization and potency matrix from iteration 2

Criterion	f_1	f_2	f_3
$d^{(2)}$	7	6	1.0
Minimal value	7	6	0.2
Maximal value	8	8	1.0

The DM decides that realization $d^{(2)}$ does not satisfy his/her expectations.

3) The DM decides that the value of criterion f_2 should be higher than 6, the value of criterion f_1 should not be lowered, but is willing to accept a value lower than 1.0 for criterion f_3:
$$\mathrm{F}_1 = \{ f_2 \}, \mathrm{F}_2 = \{ f_1 \}, \mathrm{F}_3 = \{f_3\}$$
4) Determine the set of variants which satisfy the condition formulated by the DM in step 3:
$$\mathrm{D}^{(3)} = \{d_2, d_4, d_9\}$$
5) Since $\mathrm{D}^{(3)} \neq \emptyset$, proceed to the next step.
6) Since $\mathrm{D}^{(3)}$ contains more than one realization, proceed to the next step.
7) Calculate trade-offs $t_{23}(d_i)$ for $d_i \in \mathrm{D}^{(3)}$. For each realization, $f_3(d_i) < \\ < f_3(d^{(2)})$ holds. The calculated values are:
$$t_{23}(d_2) = 3.33, t_{23}(d_4) = 3.33, t_{23}(d_9) = 0.83$$
8) Since for each realization, $f_3(d_i) < f_3(d^{(2)})$ holds, there is no need to calculate the next values of the trade-offs.
9) Trade-offs have been calculated for one pair of criteria only. For realizations d_2 and d_4 their values are identical. As the next candidate realization $d^{(3)}$ we take d_2.
10) Set $l = 3$ and proceed to the next iteration.

Iteration 3

1) Determine the potency matrix $\mathrm{M}^{(3)}$.
2) Present the criteria values obtained for realization $d^{(3)}$ and the potency matrix $\mathrm{M}^{(3)}$ (Table 7) to the DM.

Table 7: Candidate realization and potency matrix from iteration 3

Criterion	f_1	f_2	f_3
$d^{(3)}$	7	8	0.8
Minimal value	7	7	0.2
Maximal value	8	8	0.9

The DM finds realization $d^{(3)}$ satisfactory.

As the solution of the problem, realization d_2 has been finally accepted, according to which the process should begin in state 1, and the decision to be taken is 2. As a result, at the beginning of stage 2, the process will be again in state 1, and the decision to be taken at that state is 1. Eventually, the process will end in state 2. The evaluation of the selected realization with respect to criterion 1 is 6; that with respect to criterion 2 is the fuzzy number (8, 1, 1); and that with respect to criterion 3 is described by the discrete probability distribution (0.3, 0.3, 0.4).

6 Summary

The procedure presented here does not require much effort from the decision maker. When evaluating the solution proposed, all he/she must do is to divide the criteria into three groups: those whose values should be corrected, those whose values should not be made worse, and those whose values can be lowered. To determine the next candidate solution, the values of the trade-offs are analyzed.

One of the fundamental questions which should be answered when constructing a new interactive method is how to present the results to the decision maker and how the decision maker should formulate his/her preferences. This is particularly important when evaluations with respect to criteria are expressed not by real numbers but in another form. In the procedure proposed here, we scalarize the evaluations. In the future, however, we intend to propose other, more advanced tools, using other methods of interacting with the decision maker, which will allow to present him/her with more information as to the consequences of the selection of the solution.

Acknowledgements

This research was supported by National Science Centre, decision No. DEC 2013/11/B/HS4/0147.

References

Bellman R.E. (1957), *Dynamic Programming*, Princeton University Press, Princeton.

Benayoun R., De Montgolfier J., Tergny J., Laritchev O. (1971), *Linear Programming with Multiple Objective Functions: Step Method* (*STEM*), Mathematical Programming, 1(1), 366-375.

Henig M.I. (1985), *The Principle of Optimality in Dynamic Programming with Returns in Partially Ordered Sets*, Mathematics of Operations Research, 10(3), 462-470.

Korhonen P.J., Laakso J. (1986), *A Visual Interactive Method for Solving the Multiple Criteria Problem*, European Journal of Operational Research, 24(2), 277-287.

Li D., Haimes Y.Y. (1989), *Multiobjective Dynamic Programming: The State of the Art*, Control Theory and Advanced Technology, 5(4), 471-483.

Miettinen K., Makela M.M. (2000), *Interactive Multiobjective Optimization System WWW-NIMBUS on the Internet*, Computers & Operations Research, 27(7), 709-723.

Mitten L.G. (1974), *Preference Order Dynamic Programming*, Management Science, 21(1), 43-46.

Nowak M. (2006), *INSDECM – An Interactive Procedure for Discrete Stochastic Multicriteria Decision Making Problems*, European Journal of Operational Research, 175(3), 1413-1430.

Nowak M. (2007), *Aspiration Level Approach in Stochastic MCDM Problems*, European Journal of Operational Research, 177(3), 1626-1640.

Nowak M. (2010), *Trade-Off Analysis in Discrete Decision Making Problems Under Risk* [in:] Jones D., Tamiz M., Ries J. (eds), *New Developments in Multiple Objective and Goal Programming. Lecture Notes in Economics and Mathematical Systems*, Vol. 638, Springer, Berlin, Heidelberg.

Nowak M., Trzaskalik T. (2013), *Interactive Procedure for a Multiobjective Stochastic Discrete Dynamic Problem*, Journal of Global Optimization, 57(2), 315-330.

Özpeynirci Ö., Özpeynirci S., Kaya A. (2017), *An Interactive Approach for Multiple Criteria Selection Problem*, Computers & Operations Research, 78, 154-162.

Sobel M.M. (1975), *Ordinal Dynamic Programming*, Management Science, 21(9), 967-975.

Steinberg E., Parks M.S. (1979), *A Preference Order Dynamic Program for a Knapsack Problem with Stochastic Reward*, Operational Research Society Journal, 30(2), 141-147.

Steuer R.E. (1977), *An Interactive Multiple Objective Linear Programming Procedure*, TIMS Studies in the Management Sciences, 6, 225-239.

Trzaskalik T. (1991), *Wielokryterialne dyskretne programowanie dynamiczne. Teoria i zastosowania w praktyce gospodarczej*, Wydawnictwo Akademii Ekonomicznej im. Karola Adamieckiego, Katowice.

Trzaskalik T. (1998), *Multiobjective Analysis in Dynamic Environment*, Karol Adamiecki University of Economics Press, Katowice.

Trzaskalik T., Hoa D.T. (1999), *Wielokryterialne, wieloetapowe procesy decyzyjne w warunkach niepewności* [in:] Trzaskalik T. (ed.), *Modelowanie preferencji a ryzyko '99. Część II*, 409-423.

Trzaskalik T., Sitarz S. (2002), *Dynamic Discrete Programming with Partially Ordered Criteria Set* [in:] Trzaskalik T., Michnik J. (eds.), *Multiple Objective and Goal Programming. Recent Developments*, Physica-Verlag, 186-195.

Trzaskalik T., Sitarz S. (2004), *Modele programowania dynamicznego w strukturach porządkowych* [in:] Kulikowski R., Kacprzyk J. and Słowiński R. (eds.), *Badania operacyjne i systemowe '04. Podejmowanie decyzji. Podstawy metodyczne i zastosowania*, EXIT, Warszawa, 15-30.

Trzaskalik T., Sitarz S. (2007), *Discrete Dynamic Programming with Outcomes in Random Variable Structures*, European Journal of Operational Research, 177(3), 1535-1548.

Zaras K. (2004), *Rough Approximation of a Preference Relation by a Multi-Attribute Dominance for Deterministic, Stochastic and Fuzzy Decision Problems*, European Journal of Operational Research, 159(1), 196-206.