Estimating Criteria Weight Distributions in Multiple Criteria Decision Making: A Bayesian Approach

Barbaros Yet[1*], Ceren Tuncer Şakar[1]

[1]*Department of Industrial Engineering, Hacettepe University, Ankara, TR*

Corresponding Author: Barbaros Yet.

Department of Industrial Engineering, Hacettepe University, 06800, Ankara, TR.

Tel: +90 312 780 5577. E-mail: barbaros.yet@hacettepe.edu.tr

Abstract

A common way to model Decision Maker (DM) preferences in Multiple Criteria Decision Making (MCDM) problems is through the use of utility functions. The elicitation of the parameters of these functions is a major task that directly affects the validity and practical value of the decision making process. This paper proposes a novel Bayesian method that estimates the weights of criteria in linear additive utility functions by asking the DM to rank or select the best alternative in groups of decision alternatives. Our method computes the entire probability distribution of weights and utility predictions based on the DM's answers. Therefore, it enables the DM to estimate the expected value of weights and predictions, and the uncertainty regarding these values. Additionally, the proposed method can estimate the weights by asking the DM to evaluate few groups of decision alternatives since it can incorporate various types of inputs from the DM in the form of rankings, constraints and prior distributions. Our method successfully estimates criteria weights in two case studies about financial investment and university ranking decisions. Increasing the variety of inputs, such as using both ranking of decision alternatives and constraints on the importance of criteria, enables our method to compute more accurate estimations with fewer inputs from the DM.

Keywords: Multiple criteria decision making, Bayesian models, decision analysis, criteria weights, additive utility

1 Introduction

While classical optimization theory deals with problems that aim to maximize or minimize a single criterion, most realistic decision making problems have multiple conflicting criteria. Since all criteria cannot be simultaneously optimized, special methods from Multiple Criteria Decision Making (MCDM) domain are used for these problems (Belton and Stewart 2002; Greco et al. 2016). Generally speaking, a criterion is a measure of effectiveness that needs to be considered for the problem at hand, for example, profit. When the direction of improvement is added to the criterion, we obtain an objective, like maximizing profit. In general, a multiple objective problem with n maximization objectives can be formulated as in (1):

$$\text{Maximize } f(x) = \left(f_1(x), \dots, f_n(x)\right), \quad \text{subject to } x \in X \tag{1}$$

where x is the decision variable vector, X is the feasible region in decision space and f_i is the ith objective function. Let us denote the feasible region in objective space by Z. This set is the image set of X. A vector $z = (z_1, \dots, z_n) \in Z$ is called nondominated if and only if there does not exist some $y \in Z$ such that $y_i \geq z_i$ for all $i = 1, \dots, n$ and $y_i > z_i$ for at least one i. Otherwise, it is called dominated. The decision variable vector that has a nondominated objective value vector is called an efficient solution. Efficient solutions are candidate best solutions for the Decision Maker (DM) of the problem. Different DMs may have different preferences for the criteria involved in a problem, and hence may value different efficient solutions the highest. In fact, theoretically, for each of the efficient solutions involved, there can be a DM who will prefer it above the others. As a result, to determine the best solution (or decision alternative) for a DM, individual preferences have to be integrated in the decision making process. When this decision making process takes place in a continuous solution space, i.e. when X is continuous, the problem falls under the category of Multiple Criteria Optimization (MCO). On the other hand, when X is discrete, the problem is of the Multiple Criteria (Attribute) Decision Analysis (MCDA) type. MCO and MCDA are the two basic subproblems under MCDM (Steuer 1986).

In various MCDM approaches, the emphasis the DM places on different criteria is represented in the form of weights for those criteria. However, for most approaches, these weights are assumed to be known and the main effort is spent on the phases carried out with the assumed parameters. For example, the widely-used ranking method TOPSIS uses weighted distance metrics to calculate its ranking measure, but offers no mechanism to derive the weights. Two most popular outranking methods, ELECTRE and PROMETHEE, depend on the weights of criteria to determine pairwise relations of preferability among alternatives, but again work with preset weight values (see Pomerol and Barba-Romero, 2012 for descriptions of TOPSIS, ELECTRE and PROMETHEE). In a limited number of studies, approaches to estimate

preferential parameters of decision models have been proposed (see Greco et al. 2016 for a survey on such studies, some of which will be mentioned in the literature review). Eliciting and modeling the preferences of the DM is a main challenge in MCDM problems as they result in more realistic and practical solution outcomes.

Utility or value functions are widely used to represent the preferences of DMs. Although these two terms are sometimes used interchangeably, utility functions are developed for alternatives whose outcomes are uncertain whereas value functions work under certainty (Belton and Stewart 2002; Keeney and Raiffa 1993). With the certainty assumption, utility functions reduce to value functions so they are more general forms of preference. These functions aggregate the evaluations of alternatives in different criteria into an overall measure, and show the value of different solutions to the DM involved. They can be used for choosing the best alternative, choosing a set of best alternatives, ranking alternatives in preference order or sorting them into preference-ordered classes (see Greco et al. 2016 for a general review of such approaches and Doumpos and Zopounidis, 2002 for a specific review of sorting methods). The utility functions that represent the preferences of the DMs can be modeled in different forms including linear, additive, quasiconcave and general monotone. In this paper, we work in an MCDA setting where DM preferences on discrete solutions are modeled with a linear additive utility function. The DM's preferences for the criteria are expressed in the form of weights. The weights for the criteria are accepted to sum up to one, and the overall value of a solution to the DM is calculated by means of a weighted sum. Although the DM knows these weights implicitly, we realistically assume that it is difficult to express them explicitly. We propose a Bayesian method that indirectly elicits criteria weights by showing the DM small sets of decision alternatives and asking the DM to rank the alternatives in each set. We illustrate and evaluate this method by using two MCDM case studies about investment portfolio and university selection. Our method offers various benefits over previous techniques for estimating criteria weights, some of which are explained below:

1. Our method estimates the entire probability distribution of weights and utility prediction as it is a Bayesian method. This provides the DM both the expected value and uncertainty regarding the weights and utilities.
2. It is flexible in terms of the inputs required. Rather than ranking decision alternatives, the DM can also provide the best alternative or partial ranking of alternatives in each set.
3. Prior knowledge about the relative importance of weights can easily be incorporated in our method as prior distributions or constraints on weight distributions. This enables our method to estimate weights with a smaller number of evaluation inputs from the DM.

4. It asks the DM to rank elements of multiple sets each containing a small number of decision alternatives. This is an improvement over previous techniques which require the assessment of a large number of alternatives, since this brings cognitive burden on the DM and hinders consistency.

In the remainder of this paper, Section 2 reviews the relevant studies on preference elicitation, Section 3 presents our method, Section 4 illustrates and evaluates the method by using two MCDM case studies, and Section 5 concludes the paper.

2 Literature Review

In MCDM problems, preferences of the DM about the alternatives are usually expressed through utility functions of certain forms. Using the evaluations in each criterion involved, these functions make an aggregation and provide a single measure of desirability for the alternatives. Several previous approaches that employ such functions assumed that the preferential parameters of the model are known beforehand. This assumption greatly reduces the practical value of the algorithms carried out for decision making. In this section, we review studies that try to elicit DM preferences and/or account for the imprecision of the parameters of utility functions.

A classical and widely used form of utility function is the additive function. Fishburn (1967) presented an early review of methods for estimating additive utilities. He categorized several utility estimation methods with respect to factors related to the use of probabilities, type of preference judgments, the number of factors involved in judgments and the nature of solution space for the problem. One can refer to Keeney and Raiffa (1993) and Wakker (1989) for the foundations and theoretical background for additive representations of preferences. Jacquet-Lagrèze and Siskos (1982) introduced the UTA method that dwells on additive utility functions. Analyzing the preference information provided by the DM about the alternatives, the UTA method aims to find compatible additive utility functions. The DM is expected to express preferences in the form of strong preference, weak preference and indifference. Siskos and Yannacopoulos (1985) proposed an improved version of UTA, the UTASTAR method. They show that UTASTAR has better performance than the original UTA in terms of precision and the number of iterations required. The UTA method was the start of the ordinal regression paradigm in MCDM, and several forms of utility functions were considered since (see Siskos et al., 2016 for a survey). In the traditional ordinal regression-based methods, utility functions that are compatible with the decision problem are derived, and one of them is selected according to some principle. Seeing this selection as an arbitrary process, Greco et al. (2008) proposed the robust ordinal regression approach in which the whole set of utility functions compatible with the DM preferences are considered. In another approach aimed at achieving a general

assessment of alternatives, Kadzinski and Tervonen (2013) proposed a probabilistic study for robust evaluation of alternatives. Their approach calculates the probability of an alternative being preferred over another by taking the whole set of preference-compatible functions into account. Moreover, they also calculate probabilities for alternatives for achieving a certain position in the rank of preference.

Some researchers argued that some assumptions about the underlying utility functions of DMs could be restrictive and make the process of deriving a preference-compliant function infeasible. Angilella et al. (2004) discussed that the utility model assumptions of UTA may prevent the discovery of a function that is in line with DM preferences. To overcome this problem, they proposed a fuzzy integrals framework which can work with non-additive functions that are able to model interactions between criteria. Marichal and Roubens (2000) studied interacting criteria and fuzzy integrals too. They employed partial ranking information about criteria, alternatives and interactions. Benabbou et al. (2015) stated that nonlinear aggregation functions offer increased flexibility to the preference elicitation process. They used a minimax regret approach to elicit parameters of such functions.

Other studies in the literature attempted to cope with imprecise and incomplete information in the preference elicitation procedure. Salo and Hämäläinen (2001) accepted and worked with imprecise preference statements by the DM in order to reduce the effort necessary for MCDM. Sarabando and Dias (2010) worked with only ordinal information about the weights of criteria and also the values of alternatives in the criteria. They used Monte-Carlo simulation techniques to estimate the parameters of an additive aggregation model. Ahn and Park (2008) compared methods for MCDM where ordinal information (as opposed to cardinal) about weights are present. To account for the cases where the DM is unable to differentiate between certain alternatives, the issue of indifference has also been addressed in the literature. Branke et al. (2015) used indifference thresholds in robust ordinal regression. In their approach, one alternative is accepted to be preferable to another only if its value is higher by at least some threshold number. Branke et al. (2017) studied further in the area of indifference and worked on several heuristics to reduce interaction with the DM until the preferred solution is reached. See Pirlot and Vincke (2013) for the general use of thresholds in preference contexts.

Some mathematical programming approaches have also been used to estimate the weights of criteria. Pekelman and Sen (1974) proposed two models as early examples of such approaches. In their study, they first find the ideal vector of the most preferred values of the criteria. They ask the DM to make a pairwise comparison of all alternatives. Their approach asserts that the weighted distance of the preferred alternative to the ideal vector should be smaller than the distance of the other alternative to the ideal vector. Then a mathematical model is solved to find

the weights that minimize the violation of this condition in DM's responses considering all alternative pairs.

AHP, which is one of the most commonly used methods in MCDM, also includes a mechanism to elicit the weights of criteria from a DM (see Saaty, 2008 for a detailed explanation of AHP). AHP uses pairwise comparisons between the criteria and also the alternatives the DM considers. Firstly, the DM makes pairwise comparisons of criteria. The DM's degree of preference is stated in a scale between 1 (equally important) and 9 (extremely more important). Afterwards, each criterion is assigned a weight based on these comparisons. Next, the DM compares pairs of alternatives with respect to each criterion using a similar scale. This generates scores for each alternative in every criterion. Using these weights and scores, AHP computes overall scores for each alternative. AHP requires the DM to make repetitive comparisons. As the number of alternatives and criteria increase, this causes cognitive load and difficulty in making consistent preferences for the DM.

Several studies modeled utility function as a random variable and used Bayesian inference to either learn it from data or elicit it from domain experts (Chajewska et al. 2000; Chajewska and Koller 2000; Guo and Sanner 2010). Chajewska and Koller (2000) used Bayesian learning and model selection approaches on a database of partially elicited utility functions to estimate the utility distribution and structure. Chajewska et al. (2000) used value of information analysis to define the questions to elicit the utility distribution and decide when to stop. Guo and Sanner (2010) proposed an approximate preference elicitation framework and heuristics for query selection to elicit utility distributions. Bayesian models and networks have been also used for other tasks in MCDM problems than eliciting utility function parameters (e.g. see Fenton and Neil, 2001, Watthayu and Peng, 2004, Dorner et al., 2007, Sedki et al., 2010 and Delcroix et al., 2013 for MCDM applications of Bayesian models).

The focus of our study is a Bayesian framework that uses cognitively easy tasks to elicit preference information. Hence, it is flexible in terms of the form of preference information and prior knowledge used. Our framework is able to incorporate ranked or best alternative preferences from a group of alternatives, and represent prior knowledge in terms of constraints of distributions. The following section presents the details of the proposed framework.

3 Bayesian Approach

In this section, we present a novel Bayesian approach that estimates criteria weights. We focus on estimating the weights of a linear additive utility function as in (2) in this paper.

$$\sum_{i=1}^{n} w_i x_{ij} = v_j \tag{2}$$

where w_i is the weight given to criterion i, n is the number of criteria, x_{ij} is the value of criterion i for decision alternative j and v_j is the utility of decision alternative j. Keeney et al. (2006) argued that a linear additive utility function is a reasonable and accessible approach to evaluate MCDM alternatives if the criteria are appropriately chosen. Nevertheless, other types of utility functions can also be used in our method. Our aim is to elicit the values of weights w_i for a particular DM, where all weights are positive, and their sum is 1.

$$\sum_{i=1}^{n} w_i = 1 \tag{3}$$

$$w_i \geq 0 \quad i = 1, 2, \dots, n \tag{4}$$

Since direct elicitation of these weights is not viable as discussed in Section 1, we indirectly elicit them by showing the DM random groups of decision alternatives, and asking the DM to either rank the decision alternatives or select the best alternative in each group. We use a Bayesian model that updates the prior distribution of the weights based on this information.

3.1.1 Eliciting Weights by Ranking Decision Alternatives

Let $\bar{x}_j$ denote the vector of criteria values of alternative j. Suppose we show a set of k random decision alternative vectors $X = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$ to the DM. We ask the DM to rank these alternatives, and suppose the DM ranks them as in (5):

$$\bar{x}_k \succcurlyeq \bar{x}_{k-1} \succcurlyeq \cdots \succcurlyeq \bar{x}_1 \tag{5}$$

where $\bar{x}_k$ is the vector of criteria values of the best alternative, $\bar{x}_{k-1}$ is the vector of the second best alternative, and so on. Let v_j and v_l denote the comprehensive value of alternative j and l, respectively. In order to model this ranking in our model, we define a step function $u(t_{jl})$ as:

$$t_{jl} = v_j - v_l \tag{6}$$

$$u(t_{jl}) = \begin{cases} 1 & \text{if } t_{jl} \geq 0 \\ 0 & \text{if } t_{jl} < 0 \end{cases} \tag{7}$$

We need to ensure that $u(t_{jl})$ is 1 for the rankings given by the DM. For example, if the DM states that $\bar{x}_2 \succcurlyeq \bar{x}_1$, the value of $u(t_{21})$ must be 1 in the Bayesian model. This constraint is modeled by introducing a Bernoulli variable z_{21} for $u(t_{21})$ and entering 1 as an observation to this Bernoulli variable. The general form of such Bernoulli variable for $u(t_{jl})$ is shown in (8):

$$z_{jl} \sim \text{Bernoulli}(u(t_{jl})) \tag{8}$$

In order to model the DM's ranking for the set of k decision alternatives $v_k \geq v_{k-1} \geq \dots \geq v_1$, we need $k - 1$ step functions as shown by (9)-(11):

$$t_{j+1\,j} = v_{j+1} - v_j \tag{9}$$

$$u(t_{j+1\,j}) = \begin{cases} 1 \text{ if } t_{j+1\,j} \geq 0 \\ 0 \text{ if } t_{j+1\,j} < 0 \end{cases} \tag{10}$$

$$z_{j+1\,j} \sim \text{Bernoulli}(u(t_{j+1\,j})) \tag{11}$$

where $j = 1, 2, \dots, k - 1$. We call this technique the *Ranked Decision Alternatives* (RDA) approach in the remainder of the paper.
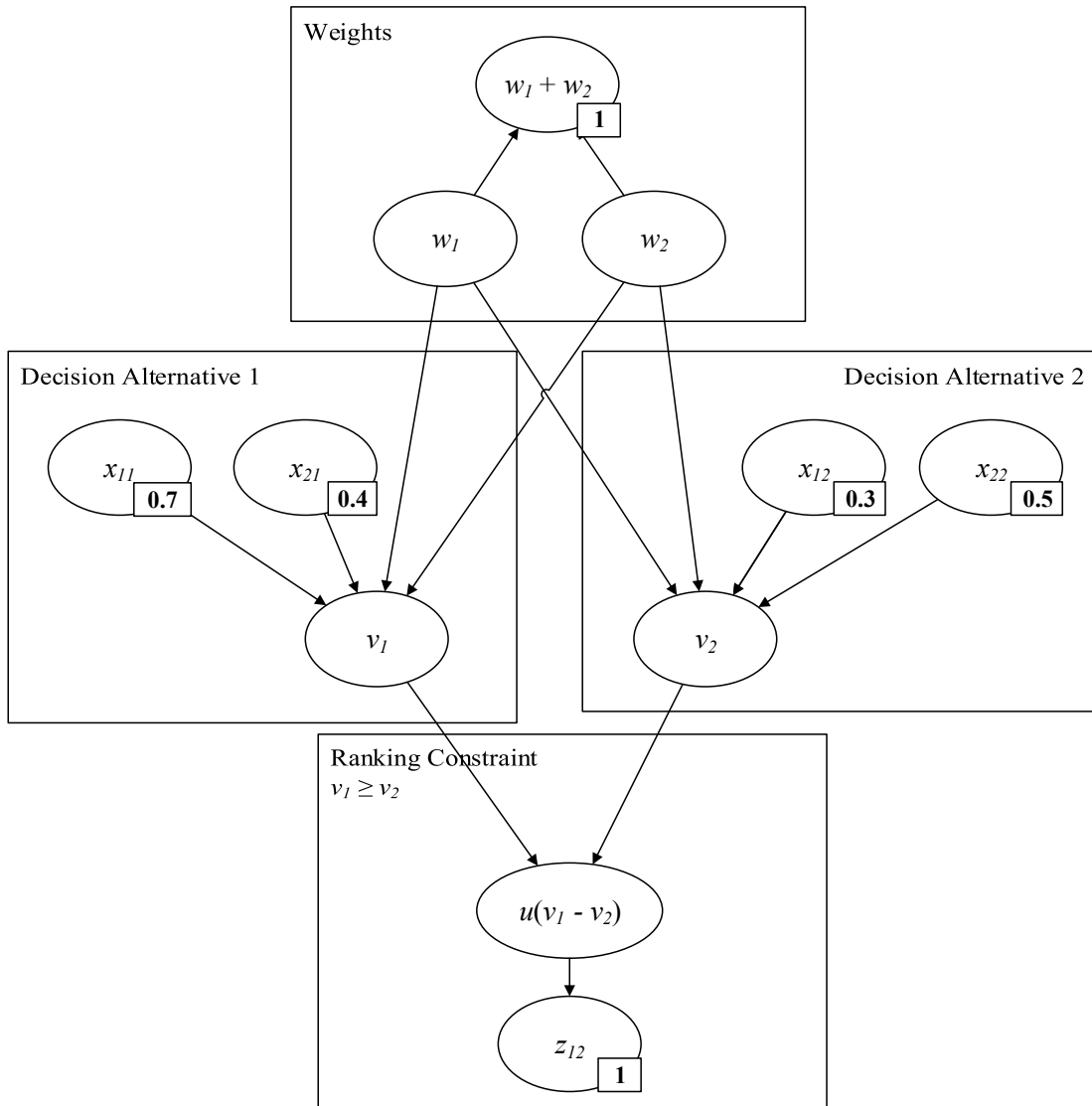

Figure 1 Bayesian Network Illustration of the RDA Model for Indirect Estimation of Weights

Figure 1 shows a Bayesian network representation of the proposed RDA approach. This example has two criteria and two decision alternatives. The variables on which we enter observations have boxes on them where the value of the observation is written. The posterior probability distributions of the rest of the variables are estimated by the model using Bayesian inference algorithms such as Markov Chain Monte Carlo (MCMC) sampling. MCMC sampling is an established and widely used inference technique for Bayesian models, and its technical details are beyond the scope of this paper (see Casella and George, 1992, Gelman et al., 2013 and Hasting, 1970 for technical details regarding MCMC).

For example, we enter the values 0.7 and 0.4 to x_{11} and x_{21} respectively before we run the model, whereas the probability distributions of w_1, w_2 and v_1 are estimated by the model. Below we describe each part of the model in Figure 1:

1. **Weights:** This part contains the probability distributions of weights w_1 and w_2 that we aim to estimate, and a constraint that ensures the sum of the weights is equal to 1. Some MCDM approaches do not require the weights to sum up to 1. In that case, the constraint can be simply removed from the Bayesian model to relax this assumption.

2. **Decision Alternative 1:** The values of the criteria and utility function for decision alternative 1 is modeled in this part. The DM knows the values of decision criteria x_{11} and x_{21}, and enters them to the model. In this example, the values of x_{11} and x_{21} are 0.7 and 0.4 respectively. We use the linear additive utility function shown by (12) for v_1. Other types of utility functions can also be used in our framework by simply changing (12). However, convergence of MCMC must be carefully assessed for different forms of functions especially when there are multimodal distributions.

$$v_1 = w_1 x_{11} + w_2 x_{12} \tag{12}$$

The decision alternative 2 has the same model structure but obviously different input values for its decision criteria. More decision alternatives can be added by replicating this model structure.

3. **Ranking $\overline{x}_1 \succcurlyeq \overline{x}_2$:** This part models the ranking provided by the DM. In this example, the DM states that the utility value of the vector of criteria values of decision alternative 1, $\overline{x}_1$, is greater than or equal to the utility value of $\overline{x}_2$. Therefore, a step function u($v_1 - v_2$) is added to the model, and the value of this step function is ensured to be 1 by entering a Bernoulli variable z_{12} as its child and entering 1 to this variable as an observation. The probability distribution of z_{12} is shown in (13):

$$z_{12} \sim \text{Bernoulli}(u(t_{12})) \tag{13}$$

After the observations about w_1+w_2, x_{11}, x_{12}, x_{21}, x_{22} and z_{12} are entered to the model, the posteriors of w_1 and w_2 can be calculated by using MCMC sampling in JAGS (Plummer 2018), Dynamic Discretization (DD) in AgenaRisk (Agena Ltd 2018; Fenton and Neil 2014) or a similar software. More criteria, decision alternatives and ranking constraints can be easily added to the model by using the structure described above. Note that, providing a larger number of rankings by the DM leads to a more precise estimation of weights. Since ranking a large number of alternatives can be difficult and time consuming for the DM, the decision alternatives can be divided into groups and each group can be ranked separately. For example, we ask the DM to rank 4 random groups of decision alternatives each with 5 elements in the case study presented in Section 4.2. Our method is flexible in terms of collecting inputs from the DM.

In the following section, we present a simpler but less precise variation of our approach that only requires the DM to select the best alternative from a group of decision alternatives, rather than ranking all the alternatives in the group.

3.1.2 Eliciting Weights by Selecting the Best Alternative

Our second approach requires the DM to only select the *Best Decision Alternative* (BDA) from a set of alternatives. Suppose we show a set of k random decision alternative vectors $X = \{\bar{x}_1, \bar{x}_2, \dots , \bar{x}_k\}$ to the DM. The DM selects the best criteria value vector $\bar{x}_m$ from X. In this case, our model must ensure that all other elements of X are equally preferred as or less preferred than $\bar{x}_m$, and revise the priors of the weights accordingly. This is modeled as in (14)-(16):

$$t_{mr} = v_m - v_r \tag{14}$$

$$u(t_{mr}) = \begin{cases} 1 \text{ if } t_{mr} \geq 0 \\ 0 \text{ if } t_{mr} < 0 \end{cases} \tag{15}$$

$$z_{mr} \sim \text{Bernoulli}(u(t_{mr})) \tag{16}$$

for each decision alternative r such that $\bar{x}_r \in X \backslash \{\bar{x}_m\}$. This model requires less effort from the DM; but we expect it to be less accurate since ranking between the other elements is not provided to the model.

3.1.3 Prior Distributions and Constraints between Weights

Our model offers the benefit of incorporating prior knowledge about the weights of the DM. We could either define informative prior distributions for the weights or introduce constraints on the values of weights. For example, if the DM believes that the weight of a decision criterion is centered on a value, we can use a Truncated Normal (TNormal) distribution with a lower bound (LB) of 0 and upper bound (UB) of 1 as shown in (17):

$$w_i \sim \text{TNormal}(\mu = 0.7, \sigma = 0.2, LB = 0, UB = 1) \tag{17}$$

If the DM believes that the weight of the criterion *i* is greater than or equal to the weight of the criterion *h*, we can introduce constraints between the weights as in (18)-(20):

$$p_{ih} = w_i - w_h \tag{18}$$

$$(p_{ih}) = \begin{cases} 1 \text{ if } p_{ih} \geq 0 \\ 0 \text{ if } p_{ih} < 0 \end{cases} \tag{19}$$

$$c_{ih} \sim \text{Bernoulli}(u(p_{ih})) \tag{20}$$

where $u(p_{ih})$ is a step function that is 1 if $w_i \geq w_h$, and it is 0 otherwise; and c_{ih} is a Bernoulli variable that ensures the posterior distribution of w_i is greater than equal to the posterior distribution of w_h when an observation value of 1 is entered on it.

Finally, if no prior information about the weights are available, uniform distributions could be used as ignorant priors:

$$w_i \sim \text{Uniform}(0,1) \tag{21}$$

3.1.4   Predictive Distributions

Our Bayesian model also offers the benefit of using the whole probability distribution of the weights when predicting the utility of a new set of decision alternatives. This enables the DM to estimate and assess the uncertainty regarding the model's weight estimations and predictions. The predictive distributions for new decision alternatives are computed as shown by (22):

$$\sum_{i=1}^{n} w_i x_{ip} = \widehat{v_p} \tag{22}$$

where w_i corresponds to the posterior weight distribution from our Bayesian model and $\widehat{v_j}$ is the predicted utility distribution of the alternative *j*. In the following section, we illustrate our Bayesian approach and evaluate its performance by using two case studies.

3.1.5   Inconsistencies

In any procedure that progresses with a DM, inconsistencies in DM responses may arise. As the process continues, the DM may give evaluations that conflict with previous ones. This can be due to the limited rationality of the human mind, an actual change of preferences on behalf of the DM or a better understanding of preferences as new alternatives become available (see Chapter 2 of French et al., 2009 for a discussion of biases and inconsistencies in decision making). Inconsistent DM responses can make the underlying models of algorithms infeasible at certain iterations. To deal with this issue, the cases that cause inconsistency can be removed

from the procedure or the degree of infeasibility caused by inconsistency can be minimized. For example, Mousseau et al. (2003) worked on a multiple criteria evaluation model where DMs provide constraints on preference parameters of the model. When the newly added constraints cause inconsistency, subsets of constraints whose removal will lead to a consistent model are identified. One can see Chinneck (2008) for different ways of handling infeasibility in algorithms.

In the proposed Bayesian approach, inconsistent answers of the DM may increase the posterior probability of weight values that are associated with those answers and were previously considered to be unlikely. In other words, inconsistent answers may lead to more uncertain and wider weight posteriors, but the inference does not necessarily stop and fail. As iterations continue, if the DM starts to provide consistent evaluations, the posterior probabilities around the true weights will start to increase again. The proposed approach can become infeasible and stop when inconsistent answers are associated with a weight that is considered to be impossible and has zero probability. This is a well-known modeling error that can be prevented by avoiding zero probabilities in priors (see Chapter 10 of Korb and Nicholson, 2010). Moreover, if the inference is done by MCMC algorithms, sufficient burn-in and sample sizes must be assigned and convergence of the algorithm must be assessed. When the DM's true utility function is non-linear or if weights are not independent, the preferences of the DM would be conceived as inconsistent for the linear utility function assumption of our approach and lead to worse predictive performance. Detecting and removing the cause of inconsistencies, however, is beyond the focus of our approach in this paper.

4   Case Studies

In this section, we present two case studies. In each case study, we made two experiments: one for the RDA approach (described in Section 3.1.1) and the other for the BDA approach (described in Section 3.1.2). In each experiment, we showed the DM multiple sets of decision alternatives and asked the DM to rank the alternatives or select the best alternative in each set. Our experiments were simulations. We used a virtual DM who ranks the alternatives or selects the best alternative according to the utility values computed by the assumed weights for the DM. We inputted this information to the RDA and BDA models to compute the posterior distributions of the weights of criteria. Note that, the true utility values were not entered to the model to estimate the weights, only the rankings given by the DM were entered.

Ranking a large number of decision alternatives could be challenging and time consuming for a DM. Therefore, we ran our experiments in multiple iterations. In each iteration, we showed the DM a set of five decision alternatives, and asked the DM to rank or select the best one only among the decision alternatives in that set. The DM ranks the alternatives in each set separately.

For example, in the second iteration, we successively enter two separate sets each having five alternatives to the model, rather than entering a single set with ten alternatives. Alternatives are evaluated with two and five criteria in our first and second case studies, respectively. Since estimating five criteria weights is a more difficult task than estimating two, it requires more preference information. Accordingly, we presented the DM two sets of alternatives in the first case study and four sets in the second. These values produced satisfactory results in our experiments; however, different values can be used as well, and the DM can also be consulted about the number of sets to evaluate.

We used JAGS to compute the posteriors of all Bayesian models in our case studies. We used the R2jags package (Su and Yajima 2015) in R as an interface to JAGS, and the mcmcplots package (Curtis et al. 2017) in R to draw diagrams. We discarded the first 15,000 samples in MCMC as the burn-in samples. We then used the autojags function from the R2jags package to compute the posteriors of the model. The autojags function updates the model in iterations until the model converges. We used $\hat{R} = 1.01$ as the convergence parameter, and updated the model with 100,000 samples at each iteration. We also assessed convergence by using sample and autocorrelation plots.

4.1 Case Study 1: Investment Portfolio Optimization

This section presents a two criteria MCDM example to illustrate the use of our approach. An investment decision is a classical MCDM example with conflicting criteria. The aim of a typical DM is to select an investment option with high return and low risk. However, the investment options with higher returns often have higher risks. Therefore, the optimal decision changes with the weights given to return and risk criteria according to the risk behavior of the DM. Some studies used more than two criteria in the portfolio problem (Koksalan and Tuncer Sakar 2016; Tuncer Sakar and Koksalan 2014). For example, some investors may want to consider multiple measures of risk such as variance, mean absolute deviation, Value at Risk and Conditional Value at Risk (CVaR), and others may be interested in additional indicators like liquidity and dividends. In this case study, we consider a portfolio problem with two criteria: expected return and CVaR, the latter being a risk measure of extreme losses. In the remainder of this section, we first describe how the investment portfolio alternatives are generated for the case study (Section 4.1.1), then we estimate the weights of a DM by using the proposed model (Section 4.1.2).

4.1.1 Generating Investment Portfolio Alternatives

We used 50 shares from the NASDAQ stock market (see Table 1) to generate investment portfolios for the DM. We collected daily prices of these stocks between August 2015 and July 2016. Using these prices, we calculated weekly returns for the stocks and computed the average

return and CVaR values from these returns. We treated the weekly return values of the stocks as equiprobable discrete scenarios. CVaR at λ probability level is the expectation of losses in the worst $(1-\lambda)$ probable cases. We used the linear model of Rockafellar and Uryasev (2000) to obtain CVaR values. Like in confidence interval calculations, 0.9 is a commonly chosen parameter for CVaR (Rockafellar and Uryasev 2000; 2002), so we have set λ to 0.9. The model is provided by (23)-(25):

$$\text{Minimize } \tau + \frac{1}{(1-\lambda)} \sum_{s=1}^{q} a_s \, p_s \tag{23}$$

$$a_s \geq -x^T r^s - \tau \qquad \forall \, s = 1, \dots, q \tag{24}$$

$$a_s \geq 0 \qquad \forall \, s = 1, \dots, q \tag{25}$$

where x is the vector of proportions of stocks in the overall portfolio, q is the number of scenarios, r^s is the vector of returns of stocks in scenario s, p_s is the probability of scenario s occurring, a_s is the auxiliary variable used for scenario s, and τ is the auxiliary variable used to calculate excess losses.

Table 1 NASDAQ Shares used in Case Study 1

NASDAQ Stock Symbol									
AAPL	DIS	F	GEVO	AXP	ARNA	AMAT	CELG	NWBO	BOFI
FB	NFLX	GILD	ETP	CRM	BA	SKLN	PTN	LUV	USAT
TSLA	GOOGL	BBRY	CHK	LC	XOM	TGRP	VG	XRX	RTN
MSFT	GOOG	SIRI	FCX	ETE	CVX	CLF	SLCA	LMT	CNC
AMZN	NVDA	DB	AA	ATVI	SDRL	SLV	SENS	ARRY	QGEN

To obtain discrete portfolio alternatives for this two criteria continuous problem, we used the augmented version of the ε-constraint method (Haimes et al. 1971). The ε-constraint method operates by optimizing one of the objective functions while the others are treated as constraints and restricted by bounds. Its general formulation for a problem where all objectives are in minimization form can be represented as:

$$\text{Minimize } f_k(x) \tag{26}$$

$$f_i(x) \leq \varepsilon_i \qquad \forall \, i \neq k \tag{27}$$

$$x \in X \tag{28}$$

where $f_i(x)$ is objective function i, ε_i is the upper bound used to constrain objective i and X is the set of feasible points. In the augmented version, the objective function is augmented by the objective functions other than k with small coefficients; this is done to ensure obtaining efficient solutions.

In our applications, we treated expected return as the constraint and divided its range into equally spaced values to generate 50 efficient portfolios. The return and CVaR values of these portfolios were then normalized between 0 and 1, where 1 is the best value for both criteria.

4.1.2   Estimating Criteria Weights

In order to evaluate our model, we made simulation experiments by using a virtual DM. Our DM is risk averse: her/his true decision weights for maximizing return is 0.4, and minimizing CVaR is 0.6. We assume that the DM has a linear additive utility function so the utility v_j of each decision alternative j is:

$$0.4 * x_{1j} + 0.6 * x_{2j} = v_j \tag{29}$$

where x_{1j} and x_{2j} are the return and CVaR values for decision alternative j, respectively. In the first case study, we showed the DM 2 sets of portfolio alternatives in each experiment.

Table 2 Two Sets of Portfolio Alternatives Ranked by the DM

Set 1			Set 2		
Return	**CVaR**	**Rank**	**Return**	**CVaR**	**Rank**
0.694	0.573	5	0.980	0.051	5
0.327	0.960	2	0.612	0.686	2
0.408	0.925	1	0.653	0.629	3
0.122	0.997	4	0.306	0.965	1
0.592	0.714	3	0.061	0.999	4

For example, Table 2 shows the return and CVaR values of 2 sets of portfolio alternatives. The DM ranked the alternatives in each set according to the linear additive utility function presented above, and these ranks are also shown in Table 2. Our method estimated the entire probability distribution of the weights and utility predictions as it is based on a Bayesian model. Figure 2 shows the probability distribution of weights computed by the RDA approach after the ranks, return and CVaR values in Table 2 are entered to the Bayesian model as described in Section 3. Our approach can predict the utility values of new portfolio alternatives after the posteriors of weights are estimated. For example, Figure 3 shows the utility distribution and 95% credible interval of a portfolio alternative predicted by using the weight distributions in Figure 2, and CVaR and return values of the portfolio. The return and CVaR values of this portfolio are 0.571 and 0.743, and the true utility value is 0.674. The expected value and 95% credible interval of this portfolio computed by our approach are 0.674 and (0.665,0.694) respectively. These probability distributions offer a more elaborate way of decision analysis as it enables us to see how confident the model is about the estimates.
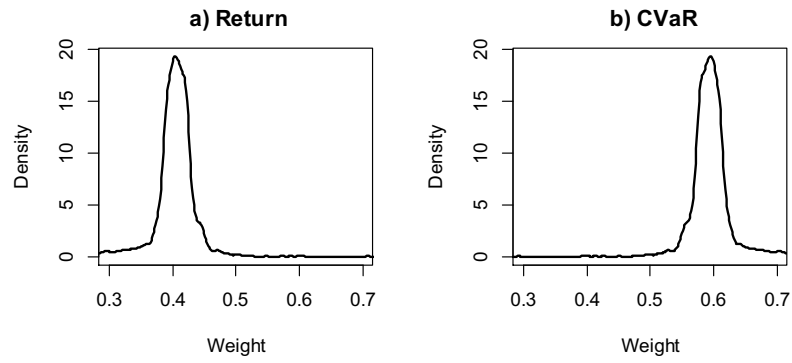
Figure 2 Posterior Probability Distributions of Weights of a) Return and b) CVaR
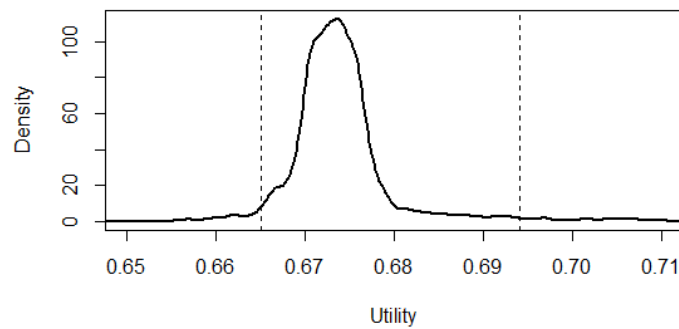

Figure 3 Predicted Utility of a Portfolio Alternative

Figures 4a and 4b show how the probability distribution of the weight of return criterion is updated after the DM uses the BDA and RDA approaches respectively. The expected value of the weights approaches to the true value and uncertainty around it decreases after every iteration. Moreover, the posteriors computed by RDA have smaller uncertainty compared to BDA. This is expected as we enter more information to the model by ranking decision alternatives than by selecting the best decision alternative.
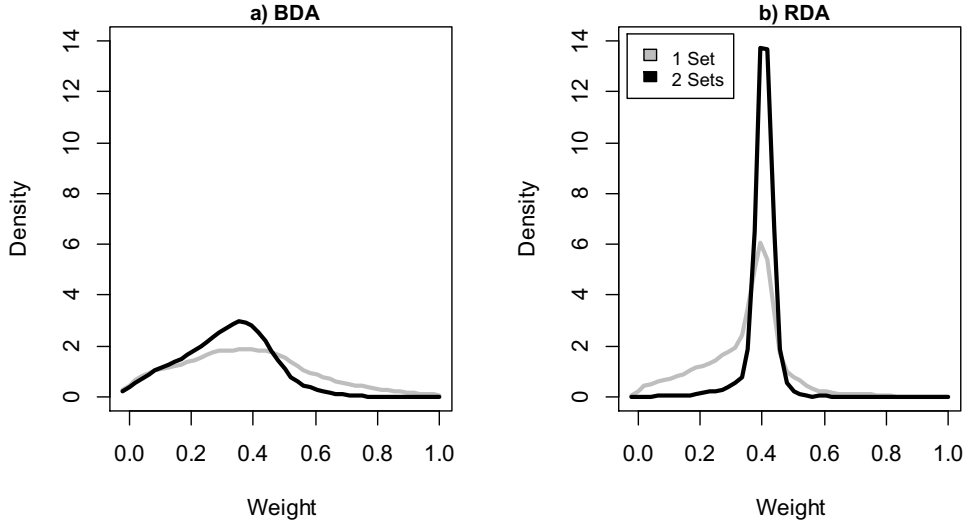
Figure 4 Posterior Probability Distribution of Weight of Return in the a) BDA b) RDA Approaches

We used 5-fold cross validation to evaluate our approach. We randomly divided the portfolios into 5 equal-sized sets. In each iteration, we estimated the weights by using one of these sets (training set), and predicted the other 4 sets (validation set) with these estimated weights. In the portfolio case study, the data contains the returns and CVaRs of 50 portfolios, therefore we estimate the weights by using 10 portfolios, and predict the value of other 40 portfolios in each iteration. Note that, we showed the training set to the DM in groups of 5 decision alternatives as described above. By the generation method (i.e. the augmented ε-constraint method), all of the decision alternatives in the dataset were non-dominated.

We used Kullback – Leibler Divergence (KLD) to measure the difference between the estimated weights and the true weights. KLD is used for measuring the difference between probability distributions. Smaller KLD values indicate similarity between the probability distributions, a KLD value of 0 shows that the estimated and the true distribution are exactly the same. Since the weights in our model have similar properties as probability distributions, i.e. the sum of the weights is 1 and each weight's value is between 0 and 1, we found KLD a suitable measure for the accuracy of weights. The KLD between the true weights P and the estimated weights Q is computed as shown by (30):

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{30}$$

Figure 5 shows how the KLD of the expected values of weights changes for the BDA and RDA approaches after each set of decision alternatives. Although the RDA approach's KLD is slightly higher when the first 5 decision alternatives are shown to the DM, it becomes smaller

17

than the BDA approach when the second 5 decision alternatives are shown. Although both approaches have similar KLD values, the RDA approach estimates the weights with less uncertainty as shown in Figure 4. Since KLD metric only uses the expected values of the weights, it disregards this additional information.
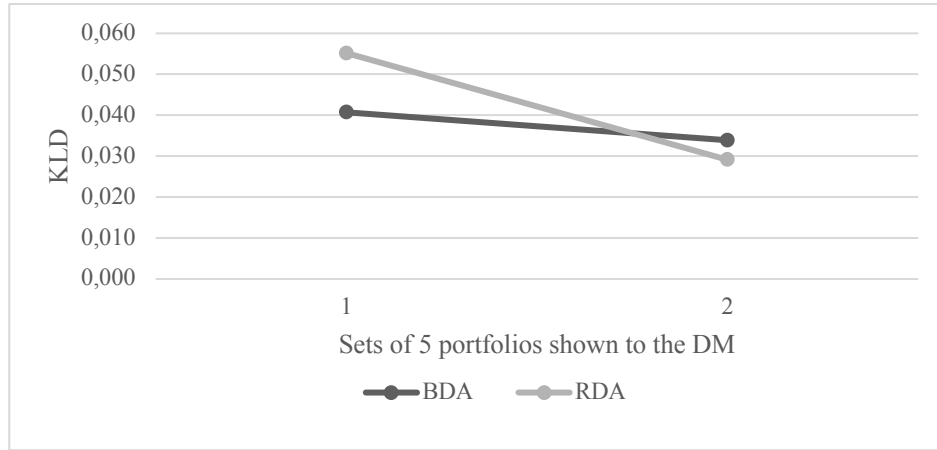

Figure 5 KLD of Estimated and True Weights in Case Study 1

In the cross validation, we predicted the utilities of the decision alternatives in the validation set. We used mean square error (MSE) to assess the predictive performance of the approach.

$$MSE = \mathrm{E}\left[\left(\mathrm{E}(\hat{v}_j) - v_j\right)^2\right] \tag{31}$$

Since MSE requires point values, we used the expected value of the predicted distribution $E(\hat{v}_j)$ for the predicted value. Therefore, MSE does not take the uncertainty around the expected values into account. Figure 6 shows the MSEs for the BDA and RDA approaches. The behavior of MSEs are similar to KLD values. In the following section, we present another MCDM case study with a higher number of criteria.
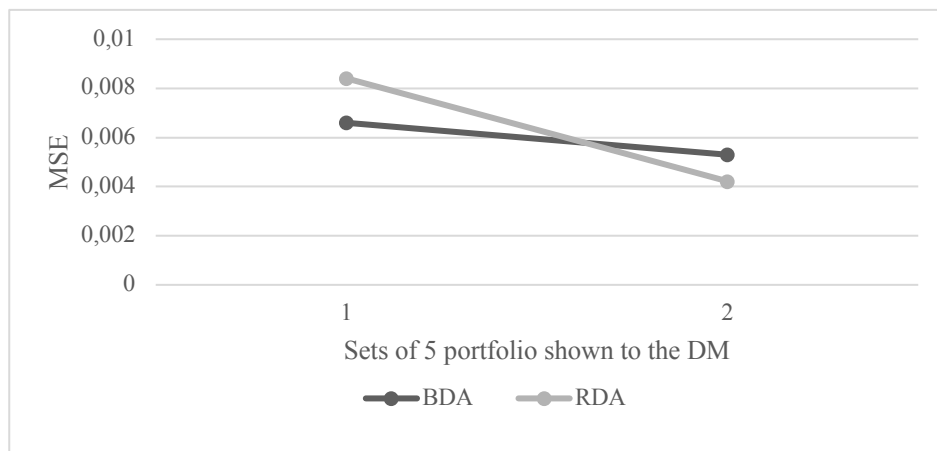

Figure 6 MSE of Utility Predictions in Case Study 1

4.2 Case Study 2: University Ranking

Times Higher Education (THE) is a journal that focuses on the news and issues about higher education and universities. Every year, THE reports university rankings for the leading universities around the world. The rankings are prepared in different categories such as world university ranking, emerging economies ranking and subject rankings. The rank of a university is defined by an overall score that is computed by a weighted average of five performance indicators. These indicators are:

- Teaching (the learning environment)
- International outlook (staff, students and research)
- Research (volume, income, and reputation)
- Citations (research influence)
- Industry income (knowledge transfer)

The weights of the performance indicators are different for different ranking categories. Table 3 shows the weights for the world university ranking category for 2015/2016 rankings. A detailed description of THE's ranking methodology is described by (Times Higher Education 2015).

Table 3 Weights of Performance Indicators

	Teaching	International Outlook	Research	Citations	Industry Income
THE World University Ranking	0.3	0.075	0.3	0.3	0.025

In our second case study, we selected 100 universities that are not dominated by each other from THE's 2015/2016 World University Rankings. In our experiments, the DM selects universities with a linear additive utility function where decision criteria and weights are the same as THE's World University Rankings:

$$0.3 * x_{1j} + 0.075 * x_{2j} + 0.3 * x_{3j} + 0.3 * x_{4j} + 0.025 * x_{5j} = v_j \qquad (32)$$

where x_{1j}, x_{2j}, x_{3j}, x_{4j}, and x_{5j} are the performance indicators of teaching, international outlook, research, citations, and industry income for university j, respectively.

Table 4 Four Sets of Universities Ranked by the DM

Teaching	Int. Outlook	Research	Citations	Ind. Income	Rank
Set1					
Teaching	**Int. Outlook**	**Research**	**Citations**	**Ind. Income**	**Rank**
0.206	0.702	0.300	0.790	0.366	4
0.355	0.587	0.206	0.840	0.329	2
0.264	0.890	0.267	0.412	0.859	5
0.354	0.744	0.439	0.665	0.395	1
0.272	0.901	0.354	0.669	0.434	3
Set2					
Teaching	**Int. Outlook**	**Research**	**Citations**	**Ind. Income**	**Rank**
0.278	0.214	0.157	0.960	0.446	3
0.329	0.320	0.454	0.342	0.969	5
0.356	0.301	0.353	0.837	0.574	1
0.323	0.485	0.270	0.813	0.823	2
0.447	0.481	0.235	0.657	0.295	4
Set3					
Teaching	**Int. Outlook**	**Research**	**Citations**	**Ind. Income**	**Rank**
0.285	0.583	0.398	0.818	0.303	1
0.610	0.442	0.558	0.071	0.285	4
0.609	0.253	0.686	0.204	0.403	2
0.453	0.293	0.427	0.494	0.747	3
0.260	0.888	0.287	0.443	0.456	5
Set 4					
Teaching	**Int. Outlook**	**Research**	**Citations**	**Ind. Income**	**Rank**
0.341	0.934	0.333	0.689	0.357	1
0.449	0.275	0.278	0.528	0.998	4
0.443	0.196	0.460	0.361	0.962	5
0.335	0.899	0.351	0.663	0.285	2
0.323	0.874	0.329	0.641	0.608	3

In the second case study, we need to estimate 5 weights and 4 of them are independent. As stated earlier, since this is a more challenging task than the portfolio case study, we showed the DM 4 sets each having 5 universities in this case study. Table 4 shows 4 sets of universities shown to the DM. We again asked the DM to rank the universities in these sets; the ranks are also shown in Table 4. Figure 7 and Figure 8 show the posteriors of weights after the information in Table 4 is entered to the model.
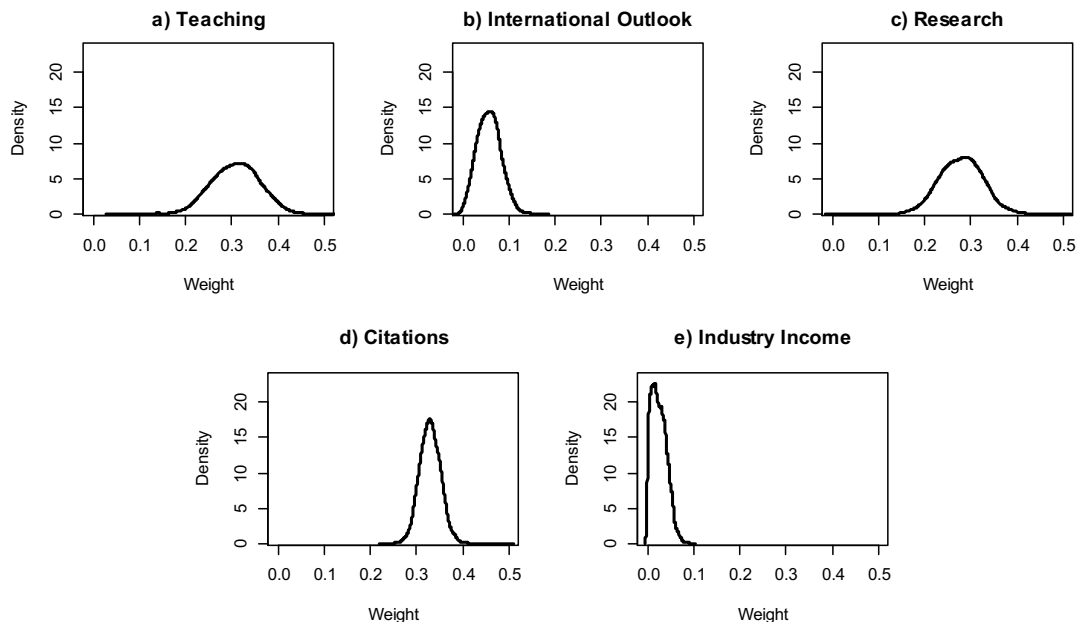
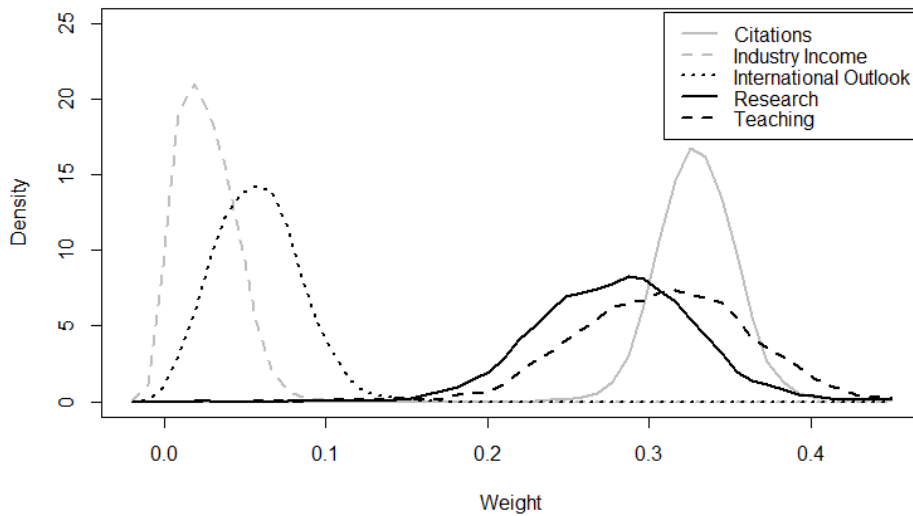*Figure 7 Posteriors of Weights in Case Study 2 Shown Separately*

Figure 8 Posteriors of Weights in Case Study 2 Shown Together

We also predicted the scores of different universities by using the probability distributions estimated by the model. For example, Figure 9 shows the predicted score of Aalborg University by using the posteriors shown in Figure 8. The true utility score of Aalborg University is 0.446 in THE, and the posterior computed by the RDA approach has an expected value of 0.452, and its 95% credible interval is (0.430,0.475).
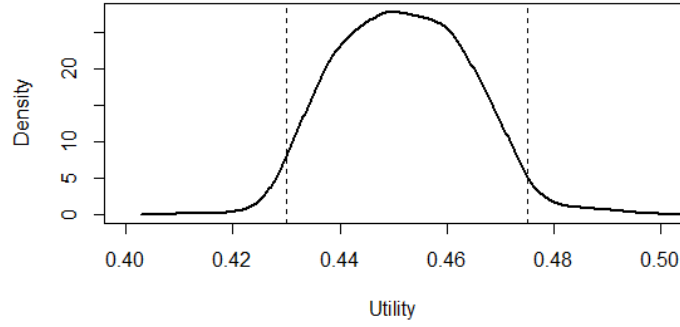
Figure 9 Predicted Utility for Aalborg University

Figure 10 shows the posteriors of the weight of teaching after the DM ranks or selects the best in each set of 5 universities in Table 4. In other words, this figure shows behavior of the proposed approach as the preference information increases. The posterior expected value of the RDA approaches the true value, and its uncertainty decreases after each set is ranked by the DM. The BDA approach has wider credible intervals, and its results do not seem to change after the third set of alternatives.
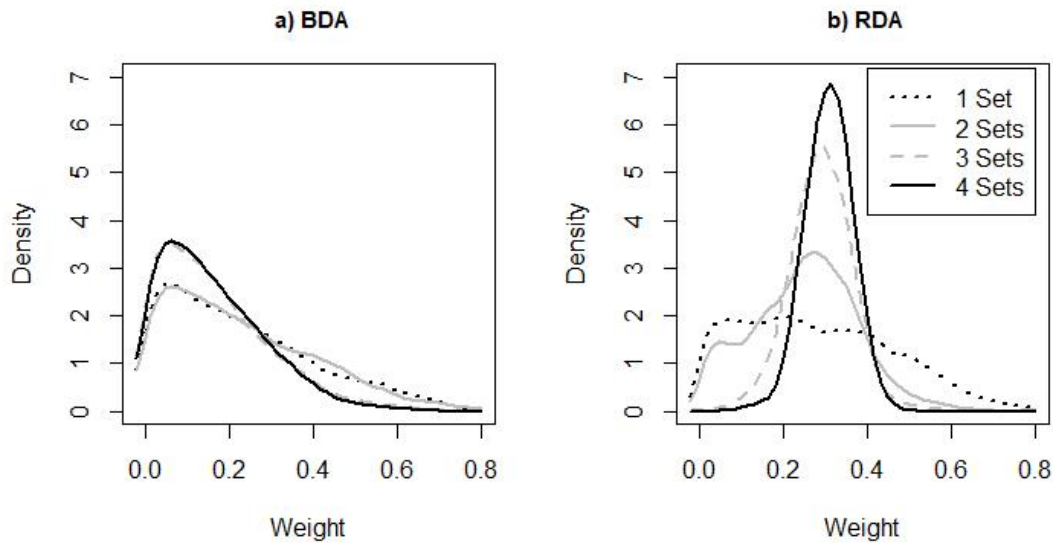
Figure 10 Posteriors of Weight of Teaching in the a) BDA b) RDA Approaches

We again used 5-fold cross validation in this case study. Our data contains 100 universities, the training sets in each iteration have 20 universities. Note that, we asked the DM to rank these 20 universities in 4 separate sets each having 5 universities. Figure 11 shows the KLD of the RDA and BDA approaches after each set of universities is shown to the DM. KLD gets smaller as more sets of universities are shown to the DM. The RDA approach consistently has lower KLD values than the BDA approach.
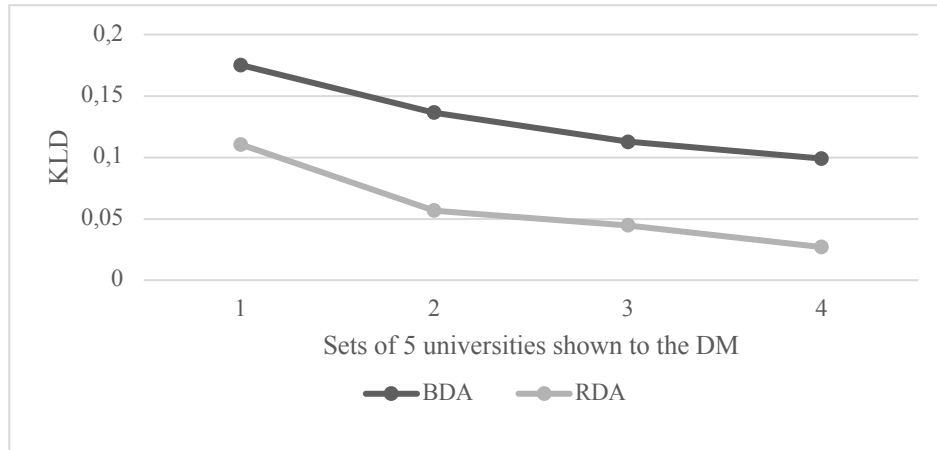
Figure 11 KLD of Weights in Case Study 2

We also used the MSEs of utility, rather than rankings, to evaluate the performance in the second case study as rankings tend to exaggerate small differences (Keeney et al. 2006). For example, if multiple universities have very similar scores, and the model makes very small errors when predicting their scores, rankings of these universities can differ in great amounts while their predicted scores do not differ significantly. MSE, however, penalizes large errors in score predictions more than small errors regardless of the similarities in the true score. Figure 12 shows the MSEs for our approaches. The MSEs of the RDA approach are consistently better than the BDA approach.
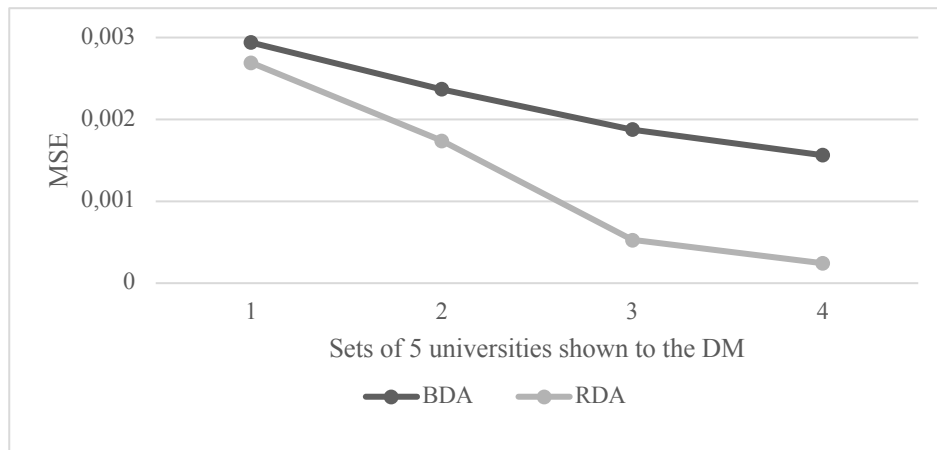

Figure 12 MSE of Utility Predictions in Case Study 2

4.3   Weight Constraints in University Case Study

If a DM is able to rank the importance of different criteria, without necessarily stating numerical values, we can also incorporate this information in our approach. In this section, we examine how adding such constraints about the importance of weights affects the performance of our model. We use the university case study for this task as it has higher number of weights that

needs to be estimated. We add the weight constraints in (33) to both the RDA and BDA approaches described in Section 3.1.3:

$$w_1 \geq w_2, \quad w_2 \geq w_5, \quad w_3 \geq w_2, \quad w_4 \geq w_2 \tag{33}$$

where w_1, w_2, w_3, w_4 and w_5 are the weights of the teaching, international outlook, research, citation and industrial income criterion, respectively. Figure 13 and Figure 14 show the posteriors of the teaching weight for the RDA and BDA approaches with and without the weight constraints. In the BDA approach, the expected value of the posterior is closer to the true value when the weight constraints are added. In the RDA approach, the weight constraints enable the model to be more confident about the value of weights even when a small number of universities are evaluated by the DM. After entering ranks of 4 sets of universities, the RDA approach computes approximately the same posterior with and without the weight constraints.
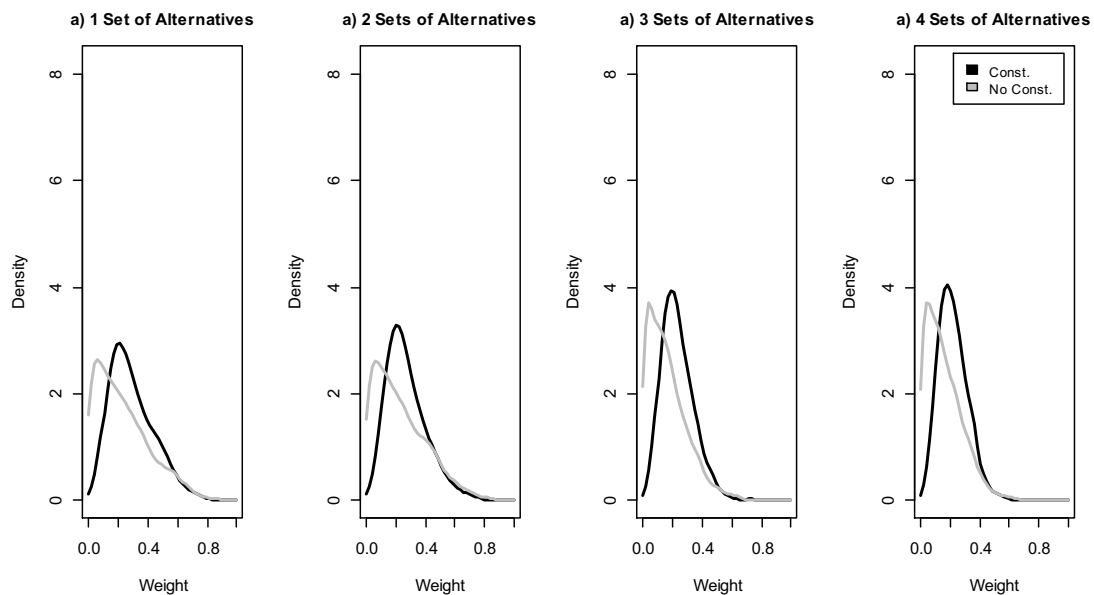

Figure 13 Posterior of Teaching with and without Weight Constraints in the BDA
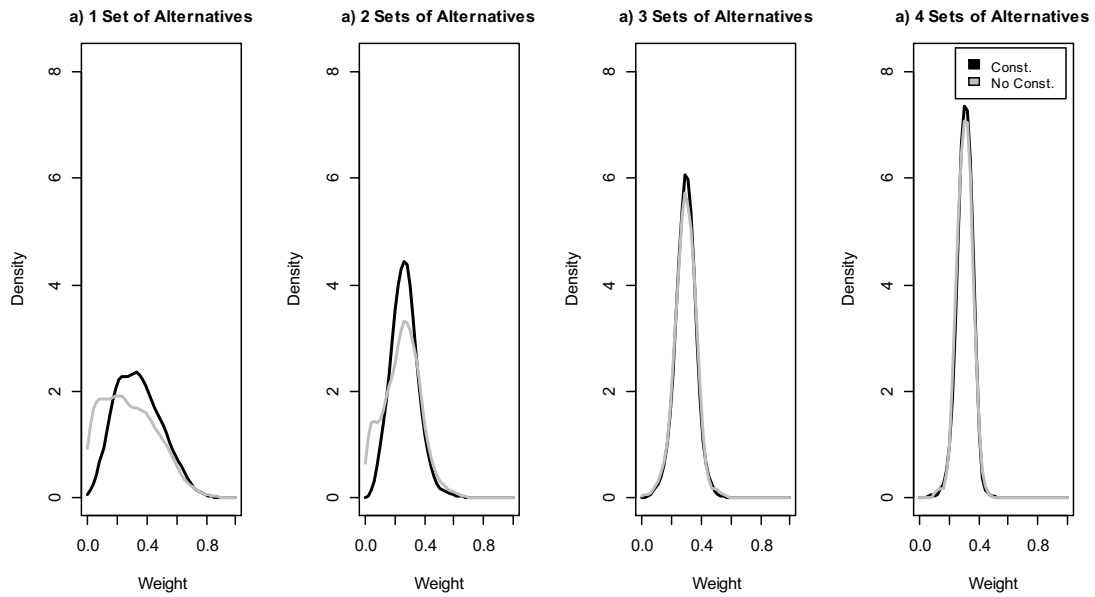
Figure 14 Posterior of Teaching with and without Weight Constraints in the RDA

Figure 15 and Figure 16 show the KLD and MSE values for these approaches. These metrics only consider the expected value of the posterior. Adding weight constraints enabled the approaches to have very close expected values to the true values, and ranking additional sets of universities did not increase the accuracy of KLD and MSE. In summary, weight constraints enabled our method to find more accurate and less uncertain results with fewer number of inputs from the DM.
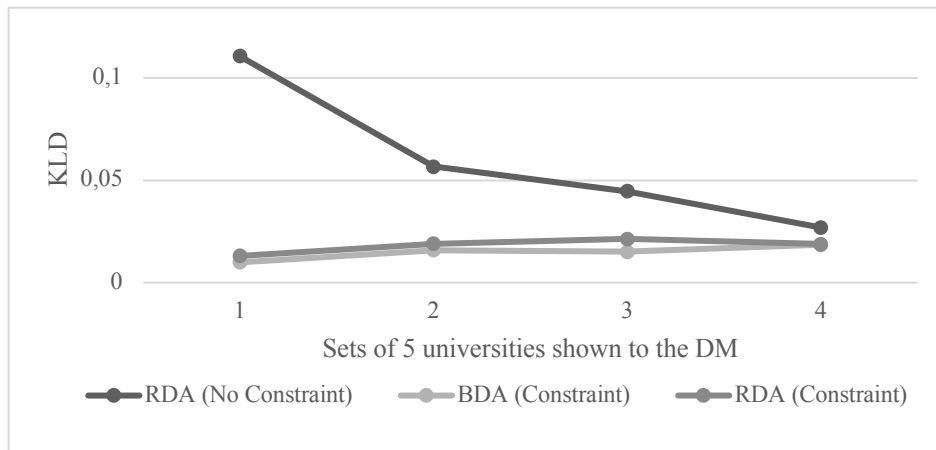

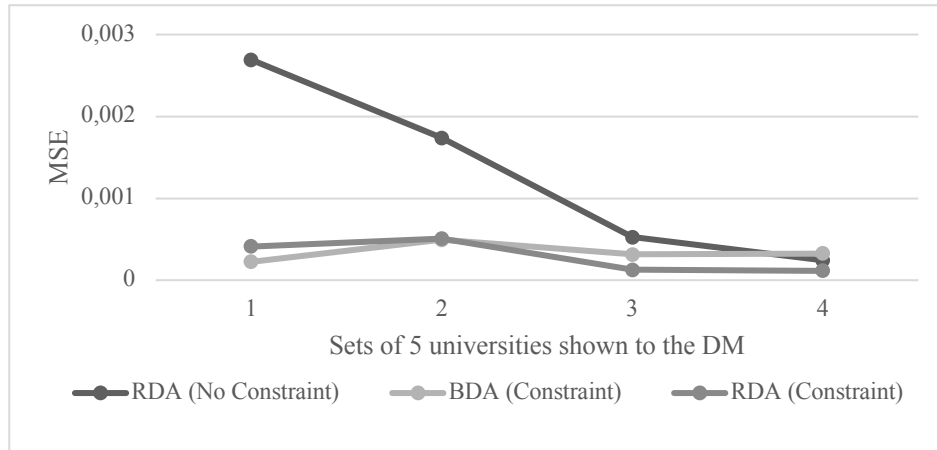*Figure 15 KLD of Weights in Case Study 2 with Weight Constraints*

Figure 16 MSE of Utility Predictions in Case Study 2 with Weight Constraints

4.4 Comparison with UTASTAR

In order to assess the performance of our method, this section compares its results to the weights inferred by the UTASTAR method by using the university ranking case study. UTASTAR is selected for comparison as it is a widely used technique for finding additive utility functions compatible with DM preferences and, its way of eliciting these preferences is similar to our method. In UTASTAR, the preferences can be stated as either a complete ranking of the alternatives shown to the DM or a set of pairwise preferences between a subset of alternatives. Although UTASTAR is not specially designed for inferring specific weights of the utility function, it can be used to infer weights by making post optimality analysis on its results (see Siskos and Yannacopoulos, 1985 and Siskos et al., 2016 for a detailed description of UTASTAR).

In our benchmark evaluation, we made 5-fold cross validation using UTASTAR. We assumed that the DM is able to identify very small differences between the utilities of alternatives and the utility function is linear as the case study is based on a linear additive utility function. We again showed the DM groups of 5 decision alternatives, and the DM ranked the alternatives in each group. In addition to this, we also used a complete ranked order of the all decision alternatives shown to the DM with UTASTAR. This second approach uses more information than RDA as ranking within groups of 5 can provide only a partial ranked order of the decision alternatives shown. However, this task is cognitively more difficult for a human DM.

Figures 17 and 18 show the KLD and MSE of RDA and UTASTAR with group ranking of alternatives and UTASTAR with complete ranking of alternatives in the university ranking case study. Using a complete ranking or ranking within each group of decision alternatives does not cause a major change in the performance of the UTASTAR approach. RDA estimates criteria weights and predicts utility scores consistently better than UTASTAR even when UTASTAR uses more information about preferences than RDA. This illustrates the ability of our method

to estimate weights by using a limited amount of information from the DM. However, it should be noted that estimation of precise weights is not the primary aim of UTASTAR.
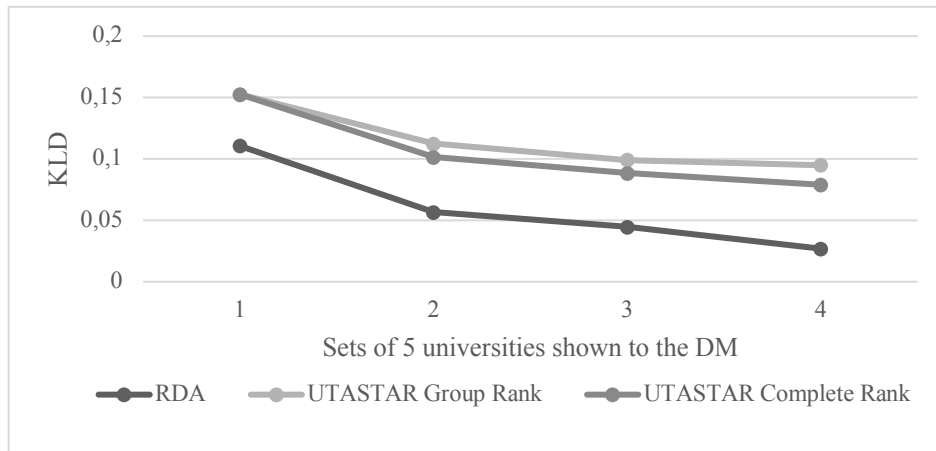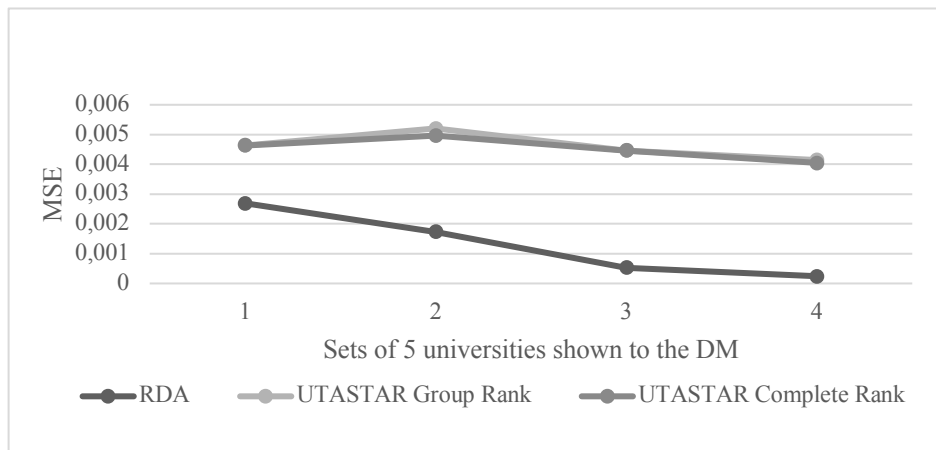

Figure 17 KLD of RDA and UTASTAR in Case Study 2

Figure 18 MSE of RDA and UTASTAR in Case Study 2

5 Conclusions

This paper proposed a novel Bayesian method that estimates the weights of decision criteria by asking a DM to select the best of or rank a set of decision alternatives. Our method offers several beneficial features for eliciting DM preferences:

- Our method estimates the entire probability distribution of weights and predicts utility values. This enables the DM to assess the uncertainty and confidence regarding these estimations.
- It can accurately estimate weights of decision criteria by asking the DM to rank multiple groups of a few decision alternatives. This is beneficial as ranking a large number of decision alternatives can be difficult and time-consuming for a DM.
- It is flexible. We proposed two variations of our method, one that ranks decision alternatives and the other that selects the best alternative, and we used a linear additive

utility function. However, other variations, such as the minimal or partial order of decision alternatives, or utility functions such as nonlinear functions, can be implemented as the Bayesian framework offers this flexibility.

- It is also able to incorporate constraints about the relative importance of weights. For example, if the DM thinks that one weight is more important than another without being able to state any numerical value, we can incorporate this information in our model to increase the accuracy of estimations.

We illustrated and evaluated the performance of our method by using two case studies. The first case study used a dataset of 50 investment portfolios evaluated by 2 criteria. The second case study used a dataset of 100 universities from THE university rankings that had 5 criteria. In both case studies, we asked the DM to select the best of or rank multiple sets each containing 5 alternatives. We entered this information to our Bayesian model and computed the posteriors of weights. We also evaluated the performance of the method by using 5-fold cross validation. In both case studies, the RDA approach had more accurate and less uncertain estimations than the BDA approach. The accuracy of the estimations increased as more number of sets were ranked by the DM. Adding weight constraints further increased the accuracy and decreased the uncertainty of the estimations. Weight constraints enabled the method to find more accurate posteriors with fewer number of inputs from the DM. We also tested our method against another preference elicitation approach, UTASTAR, and observed that our method produced superior results in terms of the accuracy of the weight estimates.

Our method asked the DM to rank random sets of non-dominated decision alternatives. As further research, we plan to identify the sets of decision alternatives that would decrease the uncertainty most drastically, and ask these questions to the DM. We expect the method to converge to the true weights much faster in this way. As another issue, in some decision problems, the weights of the DM may change over time. For example, an investor may become more risk averse as time passes. An adaptive Bayesian model can be used to identify changing weights and utility function structure. In addition, we plan to expand the method with features to detect and remove inconsistencies with DM's answers.

6 References

Agena Ltd. (2018). AgenaRisk: Bayesian Network and Simulation Software for Risk Analysis and Decision Support. http://www.agenarisk.com/. Accessed 18 January 2018

Ahn, B. S., & Park, K. S. (2008). Comparing methods for multiattribute decision making with ordinal weights. *Computers and Operations Research*, *35*(5), 1660–1670. doi:10.1016/j.cor.2006.09.026

Angilella, S., Greco, S., Lamantia, F., & Matarazzo, B. (2004). Assessing non-additive utility for multicriteria decision aid. *European Journal of Operational Research*, *158*(3), 734–744. doi:10.1016/S0377-2217(03)00388-6

Belton, V., & Stewart, T. J. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Boston: Kluwer Academic Publishers.

Benabbou, N., Gonzales, C., Perny, P., & Viappiani, P. (2015). Minimax regret approaches for preference elicitation with rank-dependent aggregators. *European Journal on Decision Processes*, 1–32. doi:10.1007/s40070-015-0040-6

Branke, J., Corrente, S., Greco, S., & Gutjahr, W. (2017). Efficient pairwise preference elicitation allowing for indifference. *Computers and Operations Research*, *88*, 175–186. doi:10.1016/j.cor.2017.06.020

Branke, J., Corrente, S., Greco, S., & Gutjahr, W. J. (2015). Using indifference information in robust ordinal regression. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9019*, 205–217. doi:10.1007/978-3-319-15892-1_14

Casella, G., & George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, *46*(3), 167. doi:10.2307/2685208

Chajewska, U., & Koller, D. (2000). Utilities as Random Variables: Density Estimation and Structure Discovery. In *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-00)* (pp. 63–71). Stanford, CA. doi:10.1.1.34.3258

Chajewska, U., Koller, D., & Parr, R. (2000). Making Rational Decisions using Adaptive Utility Elicitation. In *Proceedings of the Seventeenth National Conference on Artifical Intelligence (AAAI-00)* (pp. 363–369). Austin, TX.

Chinneck, J. W. (2008). *Feasibility and infeasibility in optimization: Algorithms and computational methods*. New York: Springer.

Curtis, S. M., Goldin, I., & Evangelou, E. (2017). mcmcplots: Create Plots from MCMC Output. https://cran.r-project.org/web/packages/mcmcplots/. Accessed 25 January 2018

Delcroix, V., Sedki, K., & Lepoutre, F. X. (2013). A Bayesian network for recurrent multi-criteria and multi-attribute decision problems: Choosing a manual wheelchair. *Expert Systems with Applications*, *40*(7), 2541–2551. doi:10.1016/j.eswa.2012.10.065

Dorner, S., Shi, J., & Swayne, D. (2007). Multi-objective modelling and decision support using a Bayesian network approximation to a non-point source pollution model. *Environmental*

Modelling and Software, *22*(2), 211–222. doi:10.1016/j.envsoft.2005.07.020

Doumpos, M., & Zopounidis, C. (2002). *Multicriteria Decision Aid Classification Methods*. Dordrecht: Kluwer Academic Publishers.

Fenton, N., & Neil, M. (2001). Making decisions: Using Bayesian nets and MCDA. *Knowledge-Based Systems*, *14*(7), 307–325. doi:10.1016/S0950-7051(00)00071-X

Fenton, N., & Neil, M. (2014). Decision Support Software for Probabilistic Risk Assessment Using Bayesian Networks. *IEEE Software*, *31*(2), 21–26. doi:10.1109/MS.2014.32

Fishburn, P. C. (1967). Methods of Estimating Additive Utilities. *Management Science*, *13*(7), 435–453. doi:10.1287/mnsc.13.7.435

French, S., Maule, J., & Papamichail, N. (2009). *Decision Behaviour, Analysis and Support*. Cambridge: Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd Editio.). Boca Raton, FL: CRC Press.

Greco, S., Ehrgott, M., & Figueira, J. R. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Berlin: Springer.

Greco, S., Mousseau, V., & Słowiński, R. (2008). Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, *191*(2), 415–435. doi:10.1016/j.ejor.2007.08.013

Guo, S., & Sanner, S. (2010). Multiattribute Bayesian preference elicitation with pairwise comparison queries. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6063 LNCS*(PART 1), 396–403. doi:10.1007/978-3-642-13278-0_51

Haimes, Y. Y., Lasdon, L. S., & Wismer, D. A. (1971). On a Bicriterion Formulation of the Problems of Integrated System Identification and System Optimization. *IEEE Journals & Magazines*, *1*(3), 296–297. doi:10.1109/TSMC.1971.4308298

Hasting, W. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*(1), 97–109.

Jacquet-Lagrèze, E., & Siskos, Y. (1982). Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. *European Journal of Operational Research*, *10*(2), 151–164. doi:10.1016/0377-2217(82)90155-2

Kadziński, M., & Tervonen, T. (2013). Robust multi-criteria ranking with additive value

models and holistic pair-wise preference statements. *European Journal of Operational Research*, *228*(1), 169–180. doi:10.1016/j.ejor.2013.01.022

Keeney, R. L., & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge: Cambridge University Press.

Keeney, R. L., See, K. E., & von Winterfeldt, D. (2006). Evaluating Academic Programs: With Applications to U.S. Graduate Decision Science Programs. *Operations Research*, *54*(5), 813–828. doi:10.1287/opre.1060.0328

Koksalan, M., & Tuncer Sakar, C. (2016). An interactive approach to stochastic programming-based portfolio optimization. *Annals of Operations Research*, *245*, 47–66. doi:10.1007/s10479-014-1719-y

Korb, K. B., & Nicholson, A. E. (2010). *Bayesian Artificial Intelligence*. Boca Raton, FL: CRC Press.

Marichal, J. L., & Roubens, M. (2000). Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, *124*(3), 641–650. doi:10.1016/S0377-2217(99)00182-4

Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., & Clímaco, J. (2003). Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research*, *147*(1), 72–93. doi:10.1016/S0377-2217(02)00233-3

Pekelman, D., & Sen, S. K. (1974). Mathematical Programming Models for the Determination of Attribute Weights. *Management Science*, *20*(8), 1217–1229. doi:10.1287/mnsc.20.8.1217

Pirlot, M., & Vincke, P. (2013). *Semiorders: Properties, Representations, Applications*. Springer Science & Business Media.

Plummer, M. (2018). JAGS: Just Another Gibbs Sampler. http://mcmc-jags.sourceforge.net/. Accessed 20 November 2018

Pomerol, J.-C., & Barba-Romero, S. (2012). *Multicriterion decision in management: principles and practice* (Vol. 25). Springer Science & Business Media.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*(3), 21–41. doi:10.2307/1165345

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, *26*(7), 1443–1471. doi:10.1016/S0378-4266(02)00271-6

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83–98. doi:10.1504/IJSSCI.2008.017590

Salo, A. A., & Hämäläinen, R. P. (2001). Preference ratios in multiattribute evaluation (PRIME)-Elicitation and decision procedures under incomplete information. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *31*(6), 533–545. doi:10.1109/3468.983411

Sarabando, P., & Dias, L. C. (2010). Simple procedures of choice in multicriteria problems without precise information about the alternatives' values. *Computers and Operations Research*, *37*(12), 2239–2247. doi:10.1016/j.cor.2010.03.014

Sedki, K., Delcroix, V., & Lepoutre, F. X. (2010). Bayesian network model for decision problems. *Intelligent Information Systems. New Approaches*, 285–298. http://iis.ipipan.waw.pl/2010/proceedings/iis10-31.pdf

Siskos, Y., Grigoroudis, E., & Matsatsinis, N. F. (2016). UTA Methods. In S. Greco, J. R. Figueira, & M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (pp. 315–362).

Siskos, Y., & Yannacopoulos, D. (1985). UTASTAR: An ordinal regression method for building additive value functions. *Investigação Operacional*, *5*(1), 39–53.

Steuer, R. E. (1986). *Multiple criteria optimization: theory, computation, and applications*. John Wiley & Sons.

Su, U., & Yajima, M. (2015). R2jags: Using R to run "JAGS." *R packages*. doi:http://cran.r-project.org/package=R2jags

Times Higher Education. (2015). World University Rankings 2015-2016 methodology. https://www.timeshighereducation.com/news/ranking-methodology-2016. Accessed 5 August 2017

Tuncer Sakar, C., & Koksalan, M. (2014). Effects of Multiple Criteria on Portfolio Optimization. *International Journal of Information Technology & Decision Making*, *13*(01), 77–99. doi:10.1142/S0219622014500047

Wakker, P. P. (1989). *Additive Representations of Preferences: A New Foundation of Decision Analysis*. Springer.

Watthayu, W., & Peng, Y. (2004). A Bayesian Network Based Framework for Multi-Criteria Decision Making. In *Proceedings of the 17th International Conference on Multiple Criteria Decision Analysis*. Whistler, British Columbia, Canada.