Universitat Pompeu Fabra Barcelona

Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1634

Accounting for estructural patterns in construction of value funcions: a convex optimization approach

Mohammad Ghaderi and Milosz Kadzińsky

January 2019

Accounting for Structural Patterns in Construction of Value Functions: a Convex Optimization Approach

Mohammad Ghaderi[1]

Miłosz Kadziński[2]

[1]Department of Business and Economics, Pompeu Fabra University, Barcelona, Spain

mohammad.ghaderi@upf.edu

[2]Institute of Computing Science, Poznań University of Technology, Poznań, Poland

Manuscript in preparation

This version: January 16, 2019

JEL Classification C44 - C13 - C53 - D90

Abstract. A common approach in decision analysis is to infer a preference model in form of a value function from the holistic decision examples. This paper introduces an analytical framework for joint estimation of preferences of a group of decision makers through uncovering structural patterns that regulate general shapes of individual value functions. We investigated the impact of incorporating information on such structural patterns governing the general shape of value functions on the preference estimation process through an extensive simulation study and analysis of real decision makers' preferences. We found that accounting for structural patterns at the group level vastly improves predictive performance of the constructed value functions at the individual level. This finding is confirmed across a wide range of decision scenarios. Moreover, improvement in the predictive performance is larger when considering the entire ranking of alternatives rather than the top choice, but it is not affected by the level of heterogeneity among the decision makers. We also found that improvement in the predictive performance in ranking problems is independent of individual characteristics of decision makers, and is larger when smaller amount of preference information is available, while for choice problems this improvement is individual-specific and invariant to the amount of input preference information.

Keywords: value function, decision analysis, convex optimization, simulation, structural patterns

1 Introduction

Shape of value function is of great importance in different areas of research in decision analysis. In multiple criteria decision making, such a shape decides upon the contribution of different performances into comprehensive value of an alternative. In decision making under risk, it captures risk attitude of a decision maker (Keeney and Raiffa, 1993). Specifically, convex or concave curvature of a value function is usually interpreted as whether the decision maker is risk seeking or risk averse, respectively, when expected utility is assumed (Fishburn and Kochenberger, 1979), and it provides information on decision maker's risk preference when prospect theory is assumed (Tversky and Kahneman, 1992; Abdellaoui et al., 2007). Moreover, S-shaped value function, proposed in the prospect theory, relates the decision maker's risk attitude to the type of outcomes (Kahneman and Tversky, 1979). In conjoint analysis, this shape describes consumers trade-off behavior (Green and Srinivasan, 1978, 1990) by revealing how they would react to the changes in product's performance levels, which, in turn, provides insights on the sensitivity of a target market to the changes in features of a product or service (Ghaderi, 2017).

The debate on shape of value function has a long history in behavioral economics and decision making (Kahneman and Tversky, 1979; Tversky and Wakker, 1995; Markowitz, 1952; Abdellaoui, 2000; Kilka and Weber, 2001; Pennings and Smidts, 2003). Based on the psychological foundations (Kahneman and Tversky, 1979), empirical evidences (Fishburn and Kochenberger, 1979), and recent insights from neuroscience (Trepel et al., 2005), several functional shapes were proposed to explain and predict consumers' choices. For instance, neoclassical economists suggested a concave shape for the value functions based on the law of diminishing marginal utility. Alternatively, Kahneman and Tversky (1979) proposed a shape that is determined by diminishing marginal sensitivity in the domains of both gains and losses. Furthermore, Markowitz (1952) considered an S-shaped value function to incorporate both reference points and loss-aversion in decision making. In addition, Friedman and Savage (1948) proposed a shape with two concave regions at the two ends of the performance scale, with a convex region between them. On the contrary, researchers in decision analysis emphasize that the decision maker's value function may hold any shape and no prior assumptions about its functional form should be made (Keeney and Raiffa, 1993; Keeney, 1996). While recognizing the merits of parametric studies, Abdellaoui (2000) argues that their findings might have been confounded by the particular parametric families chosen. Hence a disadvantage of such approach is that the estimations depend critically on the assumed functional form (Bleichrodt and Pinto, 2000).

The selection of an appropriate shape of value function is not straightforward. According to Keeney and Raiffa (1993), suitability of different techniques for assessment of value function depends on the decision problem, its context, and the decision maker's characteristics. In the context of constructing value functions, based on a series of experimental analysis and empirical research, Hershey et al. (1982) reported a significant impact of several contextual factors on the shape of value functions. Gonzalez and Wu (1999) reported a striking amount of heterogeneity in the value function's degree of curvature across individual decision makers. A relationship between the curvature of the individual value function and individual differences in preferred decision modes, i.e. intuitive versus deliberative, was found by Schunk and Betsch (2006) through an analysis of preferences of 200 students. In addition, Pennings and Smidts (2003) made a distinction between the local shape of a value function, such as local measures of curvature, and global shape of a value function, defined as the general shape of a value function over the entire outcome domain. Based on the analysis of preferences of 332 owner-managers, they reported that the global shape of value function reflects the manager's strategic decision structure, being linked to the organizational behavior (i.e., the production system employed), and is more stable than local shape that seems to drive tactical decision making (Pennings and Smidts, 2000). Similarly, Pennings and Garcia (2009) suggest a relationship between the global shape of value function and higher-order decisions. By analysis of preferences of portfolio managers, they found that the global shape, rather than the curvature, of value functions is related to the asset allocation strategies. Moreover, they suggest that the environment in which managers operate may play a role in shaping the global shape of the decision makers value function.

In this paper we aim at developing an analytical framework for constructing value functions of a group of decision makers with no prior assumptions on their shape, while acknowledging the potential existence of structural forms that would govern the general shape of these functions. The framework introduced in this paper searches for regularities in the general shape of value functions by uncovering the structural interdependencies among different regions of value functions over the performance scales or domains of outcomes. To this objective, in a unified framework, individual value functions for all decision makers are examined within each region of a performance scale and are contrasted over pairs of different regions. Comparisons over the regions are exploited to capture structural patterns that regulate the general shape of value functions. Then, these structural patterns, if any, are incorporated into the preference estimation process. Using simulation analysis and data from real decision makers, we show that accounting for structural patterns in estimation process of preferences considerably improves the predictive accuracy of constructed value functions. The underlying analytical framework is formulated in terms of convex optimization based on quadratic programming, which is computationally efficient and applicable to problems of realistic size.

2 Construction of Value Functions

Consider a set of decision makers $\mathcal{D}$ consisting of R decision makers $\{d_1, d_2, \cdots, d_r, \cdots, d_R\}$. The universe of alternatives $\mathcal{A}$ consists of N alternatives $\{a_1, a_2, \cdots, a_n, \cdots, a_N\}$. Each decision maker d_r provides a set of choice examples over a subset of alternatives $\mathcal{A}_r \subseteq \mathcal{A}$. In our framework we allow heterogeneous choice sets with different choice set size across the decision makers, and across the choice tasks for the same decision maker. We also admit indifference between profiles. These properties add a substantial flexibility in the input information, and are argued to provide several theoretical benefits in the context of choice experiment (Sándor and Wedel, 2005). In the simplest case, input information from decision makers are permitted at the minimal level, that is a set of choice examples from choice sets consisting of two profiles, where indifference between profiles is permitted. In other situations where the choice set consists of a larger number of profiles, more preference relations can be derived from the decision maker's choice examples. Alternatives are evaluated based on a family of criteria $\mathcal{G} = \{g_1, g_2, \cdots, g_m, \cdots, g_M\}$. We use the notion of "criteria" as a general concept, which can represent product attributes, decision consequences, or different points of view in assessments of decision alternatives. Each criterion $g_m \in \mathcal{G}$ evaluates each alternative according to the criterion's evaluation scale X_m, i.e. $g_m\colon \mathcal{A} \to X_m$. The evaluation scale X_m might be ratio, interval, or ordinal. Consequently, $X = \prod_m X_m$ is the evaluation space, and $x = (x_1, x_2, \cdots, x_m, \cdots, x_M)$ denotes a profile in the evaluation space.

The input preference information from a decision maker $d_r \in \mathcal{D}$ is based on holistic pairwise judgments, derived from weak preference relation $\succsim_r$ on $\mathcal{A}_r$. For a pair of alternatives $a_i, a_j \in \mathcal{A}_r$, $a_i \succsim_r a_j$ means that a_i is at least as good as a_j according to d_r's preferences. An indifference $\sim_r$ is a symmetric part of $\succsim_r$, and a strict preference relation $\succ_r$ is its asymmetric part. Therefore, each decision maker is characterized by a collection of pairwise comparisons, represented by a mixed graph $(\mathcal{A}_r, \sim_r, \succ_r)$ with the set of nodes $\mathcal{A}_r$, the set of edges $\sim_r$, and the set of arcs $\succ_r$. In special case where the indifference relation is empty, $\sim_r = \emptyset$, the mixed graph will be reduced to a directed acyclic graph (DAG).

In general, if the d_r's choice set in choice task t is defined by $\mathcal{A}_r^t \subseteq \mathcal{A}_r$, $\mathcal{A}_r^t \neq \emptyset$, and her choice involves $\mathcal{A}_r^{tc} \subseteq \mathcal{A}_r^t$, then the set of new edges derived from choice task t to be added to $\sim_r$

in the mixed graph $(\mathcal{A}_r, \sim_r, \succ_r)$ is defined by all the 2-subsets of $\mathcal{A}_r^{tc}$, and the new arcs derived from choice task t to be added to $\succ_r$ in the same mixed graph are defined by elements of the Cartesian product $\mathcal{A}_r^{tc} \times \mathcal{A}_r^t \setminus \mathcal{A}_r^{tc}$. For example, assume that $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$, and that decision maker d_r considers $\mathcal{A}_r = \{a_1, a_3, a_4, a_5\}$. Suppose that in choice task 1 she is facing the choice set $\mathcal{A}_r^1 = \{a_1, a_3, a_4\}$ and her choice involves $\mathcal{A}_r^{1c} = \{a_1, a_4\}$. From this choice example, $\{a_1, a_4\}$ (all 2-subsets of $\mathcal{A}_r^{1c}$) will be added to the edges of the mixed graph $(\mathcal{A}_r, \sim_r, \succ_r)$. In addition, all elements of $\{a_1, a_4\} \times \{a_1, a_3, a_4\} \setminus \{a_1, a_4\} = \{(a_1, a_3), (a_4, a_3)\}$ will define the new arcs in the same mixed graph. Similarly, if in the second choice task, d_r chooses $\mathcal{A}_r^{2c} = \{a_5\}$ in choice set $\mathcal{A}_r^2 = \{a_4, a_5\}$, then the mixed graph $(\mathcal{A}_r, \sim_r, \succ_r)$ will be enriched by adding the new arc (a_5, a_4) to $\succ_r$.

Preferences over the universe of alternatives are captured by a value function $V\colon X \to \mathbb{R}$. The value function V is used to construct an antisymmetric and transitive relation $\succsim$, which subsequently induces a preorder or ranking over the universe of alternatives. In this paper, we consider an additive form for the value function V, such that, for a profile $x \in X$, $V(x) = \sum_m v_m(x_m)$, where $v_m\colon X_m \to \mathbb{R}$ is the marginal value function corresponding to criterion g_m. Each marginal value function associates a numerical score with alternative's performances, hence representing their values from the perspective of a corresponding criterion. Marginal value functions are assumed to be monotonic, i.e. each criterion is of either gain- or cost-type and the corresponding marginal value function is increasing or decreasing, respectively. For a comprehensive discussion on non-monotonicity in preferences, see (Ghaderi et al., 2017; Ghaderi, 2017). Finally, we assume a piecewise linear form for the marginal value functions v_m by defining a set of breakpoints over the evaluation scales X_m. Note that additivity, monotonicity, and piecewise-linearity are the only assumptions that we make on the form of value function, without any further assumption on the structure of preferences or general shape of value function. Without loss of generality, we assume that the greater $g_m(a_n)$, the better is a_n on g_m. Also, to simplify the notion, when considering any alternative $a_n \in \mathcal{A}$, we shall write $V(a_n)$ instead of $V(g_1(a_n), \cdots, g_M(a_n))$, and $v_m(a_n)$ instead of $v_m(g_m(a_n))$, even if $V\colon X \to \mathbb{R}$ and $v_m\colon X_m \to \mathbb{R}$.

The evaluation scale X_m is divided into $\alpha_m \geq 1$ subinterval $[x_m^0, x_m^1], [x_m^1, x_m^2], \cdots, [x_m^{\alpha_m - 1}, x_m^{\alpha_m}]$ by defining $\alpha_m + 1$ breakpoints $x_m^0, x_m^1, \cdots, x_m^{\alpha_m}$ on the evaluation scale X_m. Moreover, x_m^0 and

$x_m^{\alpha_m}$ are, respectively, the minimal and maximal performances on the evaluation scale X_m. There are various ways to define the breakpoints, for instance dividing the range of performances on the evaluation scale into equal subintervals, or defining the breakpoints in such a way that each subinterval included approximately equal number of observed performances. For a comprehensive discussion on strategies for selection of breakpoints, see (Kadziński et al., 2017). In this paper, the breakpoints are defined by the entire set of distinct performance values observed over the set of alternatives; e.g. each observed performance value $g_m(a_n)$ is considered as a breakpoint.

2.1 Independent Preference Estimation Process

In this section, we present a Linear Programming (LP) technique for constructing value function for a single decision maker, d_r, from her choice examples. For a decision maker d_r with pairwise comparisons $(\mathcal{A}_r, \sim_r, \succ_r)$, a value function V^r composed of a set of marginal value functions $v_m^r \colon X_m \to \big[0, 1\big]$ is constructed, so that the preference information provided by the decision maker is reproduced. For a decision maker d_r, an unknown parameter v_m^{rj} is associated to each breakpoint x_m^j, where $v_m^{rj} = v_m^r(x_m^j) - v_m^r(x_m^{j-1})$, $1 \leq j \leq \alpha_m$. Therefore, the marginal value for d_r at a breakpoint x_m^ℓ is obtained as $v_m^r(x_m^\ell) = \sum_{j=1}^{\ell} v_m^{rj}$. Note that the unknown parameters v_m^{rj} are defined as the difference between marginal values of two consecutive breakpoints, hence linked to the slope of marginal value function. Such definition of unknown parameters has computational, theoretical, and practical advantages. First, it reduces the size of the optimization problem substantially by reducing number of constraints through converting the monotonicity constraints to sign constraints on the parameters. Second, while it is hard to interpret absolute level of marginal value function at a particular point, the change in marginal values is directly associated to the customer's source sensitivity and provides information on the trade-off analysis in the customer's assessments of alternatives. Third, identification of structural patterns in our framework is based on such source sensitivities, hence definition of unknown parameter in the current form fulfills the task more conveniently. For a d_r, the mixed graph $(\mathcal{A}_r, \sim_r, \succ_r)$ is translated into a set of linear constraints:

$$\left.\begin{array}{ll} \text{for } a_i, a_j \in \mathcal{A}_r : & \\ \quad V^r(a_i) - V^r(a_j) \geq \varepsilon_r, & \text{if } (a_i, a_j) \in \succ_r \\ \quad V^r(a_i) - V^r(a_j) = 0, & \text{if } \{a_i, a_j\} \in \sim_r \end{array}\right\} \; E_{d_r}^{\mathcal{A}_r} \tag{1}$$

where ε_r is a positive parameter used to discriminate between comprehensive values of alternatives in $\succ_r$. The set of all value functions $\mathcal{U}_r$ compatible with $(\mathcal{A}_r, \sim_r, \succ_r)$ is defined by the following set of constraints:

$$\left.\begin{array}{ll} E_{d_r}^{A_r}, & \\ v_m^r(x_m^0) = 0, & m = 1, \cdots, M, \\ \sum_{m=1}^{M} v_m^r(x_m^{\alpha_m}) = 1, & \\ v_m^{rj} \geq 0, \; j = 1, \cdots, \alpha_m, & m = 1, \cdots, M, \end{array}\right\} \; E(\mathcal{U}_r) \tag{2}$$

with the last three constraints defining the conditions on normalization and monotonicity of value functions. In order to obtain the most discriminant value function in $\mathcal{U}_r$ representing the preferences of $d_r \in \mathcal{D}$, the following LP problem has to be solved:

$$Maximize \; \varepsilon_r, \quad \text{subject to} \quad E(\mathcal{U}_r). \tag{3}$$

Let us denote by ε_r^* the optimal value of ε_r by solving the above LP problem. There exists at least one value function compatible with $(\mathcal{A}_r, \sim_r, \succ_r)$, if $E(\mathcal{U}_r)$ is feasible and $\varepsilon_r^* > 0$. Otherwise, some comparisons of reference alternatives cannot be reproduced. The reason for such an incompatibility could be in the assumptions made about the preference model (i.e., monotonic and additive value function) or inconsistency between the decision maker's judgments. In decision aid practices where an interaction with the decision maker is admissible and desirable, she would be asked to revise some comparisons in order to obtain a compatible value function, or to pursue the analysis while accepting some level of incompatibility. In situations where interaction with the decision maker is not possible, other techniques (e.g., identifying a minimal set of comparisons to be dropped) might be pursued. For a detailed discussion on dealing with incompatibility, see (Greco et al., 2008).

3 Proposed Framework: Accounting for Structural Patterns in Construction of Value Functions

The proposed framework aims at constructing value functions for a group of decision makers with no prior assumption on their shapes, while accounting for the regularities observed in the general shape of these functions captured with some structural rules governing their curvature and global shape. The framework is based on identifying the interdependencies among different regions of a value function by joint evaluation of individual value functions across a group of decision makers.

3.1 Theoretical Framework

In the proposed framework, we base our analysis on the slope of value functions, i.e., sensitivity to the decision output or *source sensitivity*, which controls curvature of a value function. For each region of the performance scale on a given criterion and for each individual decision maker, we compare deviation of the individual source sensitivity from the group's mean. deviation of d_r's source sensitivity from the group's mean is defined as the individual's *relative sensitivity*, that can be positive or negative. Then we examine the interdependencies between different regions of a value function in terms of relative sensitivities. Our objective is to identify the patterns governing curvature and global shape of a value function over some performance sub-regions or the entire domain of performance scale. To this aim, all regions of value function are compared pairwise in terms of individual relative sensitivities. Comparing marginal values over two adjacent subintervals captures concavity of the constructed value function at the shared breakpoint of the subintervals, therefore, accounting for the curvature. However, by comparing relative sensitivities across all possible pairs of subintervals, rather than merely the adjacent ones, our approach accounts for the general shape of value functions.

Note that identification of the structural patterns is not based on a particular family of value functions, neither similarities in utility levels (values), source sensitivities (slopes), curvatures, or general shapes. Instead, regularities are explored at a higher level of abstraction by investigating interdependencies among deviations of source sensitivities from the group's mean over the entire region of evaluation scale across all the decision makers. From this perspective, this approach is "borrowing" information at the group-level to incorporate in the preference estimation at the individual-level, while capturing heterogeneity in preferences at a maximal level.

To clarify this idea, let us consider an illustrative example with a group consisting of five decision makers with three different general shapes in their value functions: concave, convex, and linear (see Figure 1). By examining their value functions, a comparison between regions A and C implies that the decision makers with a positive relative sensitivity (greater source sensitivity relative to the group's mean) over region A tend to demonstrate a negative relative sensitivity (smaller source sensitivity relative to the group's mean) over region C. This observation is confirmed for the two decision makers with concave value functions. Moreover, the decision makers with a negative relative sensitivity over region A tend to demonstrate a positive relative sensitivity over region C, an observation confirmed by the two decision makers with convex value functions. The two observations together reveal a negative association between regions A and C in terms of relative sensitivities. Such information can be incorporated in the estimation process regardless of curvature or general shape of value function for a decision maker. Furthermore, the decision makers demonstrate roughly similar level of source sensitivity in region B. In the proposed framework, this is captured by adjusting the source sensitivities with reference to the group's mean, depending on the variation level in the relative sensitivities across the entire group.

In the above example, comparisons are made over three regions of the performance scale for value functions of standard forms. However, similar arguments can be extended to value functions with a more complex shape. The proposed framework can capture the structural forms regardless of the shape of value function or level of heterogeneity in value functions' shapes across the decision makers. This is due to comparing all $\alpha_m \times (\alpha_m - 1)/2$ pairs of regions over the performance scale of each criterion g^m, seeking for consistent and systematic interdependencies among these regions in terms of relative sensitivities. Therefore, effectiveness of the proposed framework is not influenced by the level of complexities in functional form of value functions.

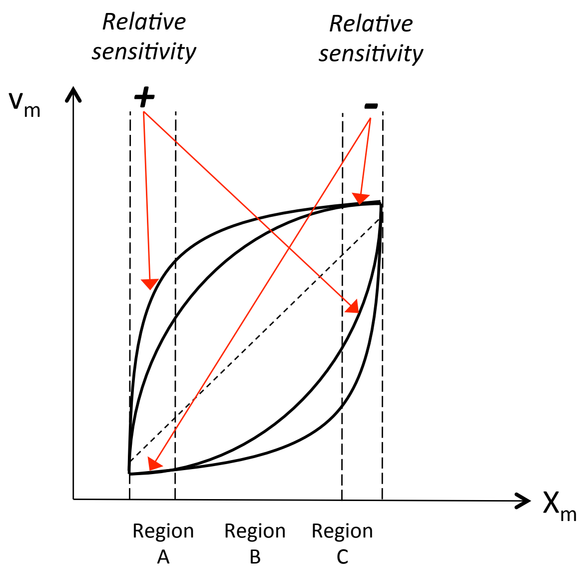

Figure 1: Interdependencies among the regions of value functions in the presence of heterogeneous decision makers.

3.2 Analytical Framework: Joint Preference Estimation Process

The proposed analytical framework estimates preferences of all decision makers jointly and simultaneously by exploiting their preference structures and accounting for the structural patterns in the preference models. Suppose that v_m^{rj} , $j \in \{1, \cdots, \alpha_m\}$, is a gain in the marginal value when improving the performance from x_m^{j-1} to x_m^j for decision maker $d_r \in \mathcal{D}$ over the j^{th} subinterval on evaluation scale X_m. Further, $\overline{v}_m^{\cdot j}$ is the average of gains in the marginal values over the j^{th} subinterval of X_m for all decision makers in $\mathcal{D}$. Define σ_m^j as the empirical standard deviation of $v_m^{rj} - \overline{v}_m^{\cdot j}$ values. Term σ_m^j represents the heterogeneity in relative sensitivities. Moreover, define ρ_m^{jk} as the empirical Pearson correlation coefficient between the relative sensitivities in the j^{th} and k^{th} subintervals. Term $\rho_m^{jk} \in [-1, 1]$ denotes the level of association between relative sensitivities over the j^{th} and k^{th} subintervals on X_m.

A large positive value of ρ_m^{jk} implies that individual's relative sensitivities tend to be in the same direction over the j^{th} and k^{th} subintervals, while a negative ρ_m^{jk} means the opposite. In other words, a positive value of ρ_m^{jk} implies that a decision maker with a greater (smaller), relative to the group, source sensitivity over the j^{th} subinterval tends to demonstrate a greater (smaller) source sensitivity over the k^{th} subinterval too. Similarly, a negative value of ρ_m^{jk} implies that the decision maker's attitudes to the performance changes over the j^{th} and k^{th} subintervals are, relative to the group, opposite.

To operationalize this idea, the following two objectives are considered simultaneously:

$$\begin{aligned}
&i)\ \rho_m^{jk}(v_m^{rj} - \overline{v}_m^{\cdot j})(v_m^{rk} - \overline{v}_m^{\cdot k}),\\
&ii)\ (1 - \frac{{\sigma_m^j}^2}{\sum_{\ell=1}^{\alpha_m} {\sigma_m^\ell}^2})(v_m^{rj} - \overline{v}_m^{\cdot j})^2,\\
&\forall j,k \in \{1,\cdots,\alpha_m\}, m \in \{1,\cdots,M\}, r \in \{1,\cdots,R\}.
\end{aligned} \tag{4}$$

The first term, to be maximized, accounts for the interdependencies between two subintervals in terms of relative sensitivities. A positive (negative) correlation between two subintervals implies the reinforcement of slope deviations form the group's mean over the two subintervals in the same (opposite) direction. The intensity of this reinforcement depends on the correlation's strength.

The other term, to be minimized, readjusts the marginal value of $d_r \in \mathcal{D}$ over a subinterval by accounting for variation in relative sensitivities across the group. The level of variation in a given subinterval is standardized with respect to the total variation in relative sensitivities over the entire performance scale. The higher the variation level, the weaker a "push" to the group's mean.

The two terms, considered together, uncover the structural forms in value functions by accounting for the interdependencies among pairs of subintervals and the variations within each subinterval. Consequently, a value function for each decision maker is obtained by solving the following convex optimization problem:

$$\begin{aligned}
Maximize\ \ & \frac{1}{\eta}\frac{1}{\kappa_1^*}\sum_r \varepsilon_r + \frac{1}{\kappa_2^*}\Big(\sum_{m,j,k}\rho_m^{jk}\sum_r(v_m^{rj} - \frac{1}{R}\sum_r v_m^{rj})(v_m^{rk} - \frac{1}{R}\sum_r v_m^{rk})\\
& - \sum_{m,j}(1 - \frac{{\sigma_m^j}^2}{\sum_{\ell=1}^{\alpha_m}{\sigma_m^\ell}^2})\sum_r(v_m^{rj} - \frac{1}{R}\sum_r v_m^{rj})^2\Big)\\
s.t.\ \ & E(\mathcal{U}_r), E_{d_r}^{\mathcal{A}_r},\\
& \quad \varepsilon_r \geq t_r, \forall r \in \{1,\cdots,R\},
\end{aligned} \tag{5}$$

where $0 < t_r \leq \varepsilon_r^*$ is a predefined small positive value, ensuring a minimal level of discrimination between comprehensive values of pairs of alternatives belonging to the strict preference relation in the constructed value function for each decision maker (with ε_r^* being the maximal value of parameter ε_r according to the choice examples of $d_r \in \mathcal{D}$), and is set to 0.001 in all the subsequent analysis in this paper. The exogenous parameter η determines an extent to which the uncovered structural patterns at the group-level are incorporated in the preference estimation process at the individual-level. Constructed value functions converge to those derived from independent estimation of value functions when η approaches zero.

Moreover, parameters ρ_m^{jk} and σ_m^j indicate the interdependencies between the regions and variation within regions, respectively, in terms of relative sensitivities. Their values are estimated from the individually and independently inferred value functions for each decision maker. The values of variables v_m^{rj} and ε_r are to be determined. Finally, the following two constants $\frac{1}{\kappa_1^*}$ and $\frac{1}{\kappa_2^*}$ are scale factors. Constant κ_1^* is equal to $\sum_r \varepsilon_r^*$, whereas κ_2^* is computed by plugging in the estimated v_m^{rj} values from the independent estimation of preferences into the second and third terms of the objective function. In addition to the computational benefits, these scale factors make the three terms in the objective function unit-less and comparable in magnitude, thus making the interpretation of parameter η independent from the particular decision setting. Let us emphasize that the proposed analytical framework can be applied to both qualitative and quantitative criteria as it builds on the associations between the deviations in marginal values rather than the original performances of alternatives.

3.3 Illustrative Example

Let us consider an example decision problem involving $R = 4$ decision makers and $N = 4$ alternatives evaluated in terms of $M = 2$ quantitative criteria. Note that in our framework criteria can also be qualitative with an ordinal evaluation scale. Nevertheless, for simplicity in description, we consider only quantitative criteria in this illustrative example. The performances of alternatives over the two criteria are provided in Table 1. Assume that the decision makers evaluate the alternatives according to their individual value functions which are presented in Table 2, hence obtaining different rankings, as presented in the same table.

Suppose that each decision maker provides incomplete preference information in form of pairwise comparisons presented in Figure 2. When analyzing preferences of each decision maker

Table 1: Performances of alternatives on two criteria considered in the illustrative decision problem.

	g_1	g_2
a_1	0.8	0.3
a_2	0.0	1.0
a_3	1.0	0.0
a_4	0.3	0.8

Table 2: Value functions and respective rankings of alternatives for the four decision makers.

Decision Maker	Value function	Ranking
d_1	$V^1(x_1, x_2) = 0.6x_1^2 + 0.4x_2$	$a_3 \succ a_1 \succ a_2 \succ a_4$
d_2	$V^2(x_1, x_2) = 0.35x_1^2 + 0.65(1 - (1 - x_2)^2)$	$a_4 \succ a_2 \succ a_1 \succ a_3$
d_3	$V^3(x_1, x_2) = 0.4x_1 + 0.6(1 - (1 - x_2)^2)$	$a_4 \succ a_1 \succ a_2 \succ a_3$
d_4	$V^4(x_1, x_2) = 0.5(1 - (1 - x_1)^2) + 0.5x_2^2$	$a_4 \succ a_1 \succ a_2 \sim a_3$

independently with the standard preference disaggregation approach in Section 2.1, we obtained their individual value functions as presented in Figure 3.

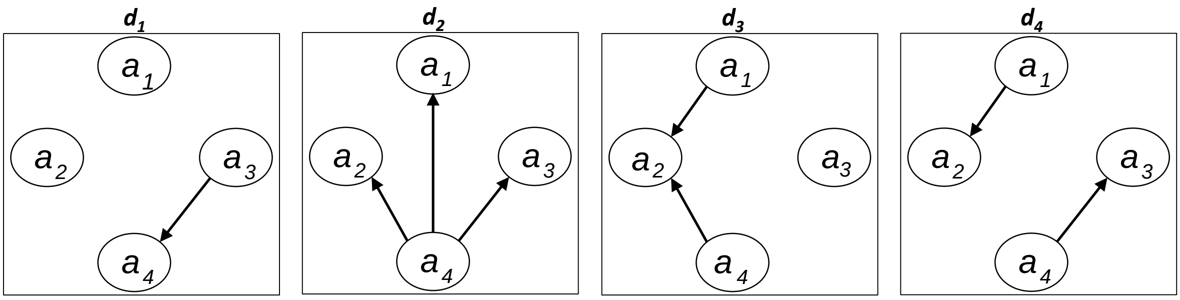

Figure 2: Pairwise comparisons provided by the four decision makers.

In our method, we aim at discovering structural patterns in the general shape of value functions by comparing all pairs of the subintervals in performance scale of each criterion in terms of slope deviations from the group's mean, and by joint consideration of the entire set of decision makers. By focusing on g_1 and comparing the marginal values of individual decision

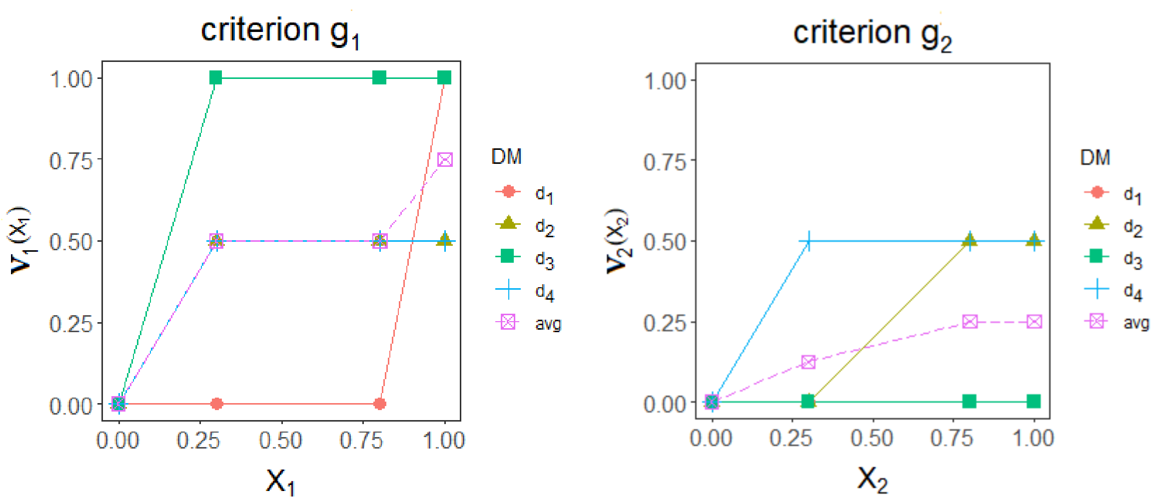

Figure 3: Constructed value functions with the standard preference disaggregation approach for the four decision makers.

makers with the group's average (dashed line in Figure 3), a negative correlation between the first and third subintervals is observed. That is, when for a value function the slope over the first subinterval is greater (lesser) than the average one, the slope over the third subinterval is lesser (greater) than the average slope. The marginal values over each subinterval for each decision maker and the average values, obtained from the independent estimation of preferences, are presented in Table 3.

Table 3: Changes in the marginal values over each subinterval of X_1 and X_2 for the four decision makers.

Criterion g_1						Criterion g_2					
Subinterval	d_1	d_2	d_3	d_4	Average slope	Subinterval	d_1	d_2	d_3	d_4	Average slope
[0.0, 0.3]	0.0	0.5	1.0	0.5	0.5	[0.0, 0.3]	0.0	0.0	0.0	0.5	0.125
[0.3, 0.8]	0.0	0.0	0.0	0.0	0.0	[0.3, 0.8]	0.0	0.5	0.0	0.0	0.125
[0.8, 1.0]	1.0	0.0	0.0	0.0	0.25	[0.8, 1.0]	0.0	0.0	0.0	0.0	0.0

By comparing individual source sensitivities in each subinterval with the group's average, we can compute the variances of relative sensitivities in each subinterval and correlations between pairs of subintervals. For example, the correlation between the first and third subintervals of X_1 can be calculated as follows:

$$\rho_1^{13} = \frac{covariance([0-0.5, 0.5-0.5, 1-0.5, 0.5-0.5], [1-0.25, 0-0.25, 0-0.25, 0-0.25])}{\sqrt{variance([0-0.5, 0.5-0.5, 1-0.5, 0.5-0.5]) \cdot variance([1-0.25, 0-0.25, 0-0.25, 0-0.25])}} =$$

$$= \frac{\frac{1}{3}[(-0.5)\cdot .75 + 0.5\cdot(-0.25)]}{\sqrt{\frac{1}{3}[(-0.5)^2 + 0.5^2]\cdot\frac{1}{3}[0.75^2 + (-0.25)^2 + (-0.25)^2 + (-0.25)^2]}} = -0.82.$$

The computed variances and correlations for all subintervals are presented in Table 4.

Table 4: Variances at each subinterval and correlations between different pairs of subintervals for X_1 and X_2.

	Criterion g_1				Criterion g_2		
Subinterval	[0.0, 0.3]	[0.3, 0.8]	[0.8, 1.0]	Subinterval	[0.0, 0.3]	[0.3, 0.8]	[0.8, 1.0]
[0.0, 0.3]	0.6			[0.0, 0.3]	0.5		
[0.3, 0.8]	0.0	1.0		[0.3, 0.8]	0.33	0.5	
[0.8, 1.0]	-0.82	0.0	0.4	[0.8, 1.0]	0.0	0.0	1.0

These measures are incorporated in the σ_m^j and ρ_m^{jk} values in the convex optimization problem (5). Moreover, the scaling factors κ_1^* and κ_2^* are computed from the independent estimation of preferences as 3 and 1.52, respectively. For the considered example, the optimization problem takes the following form:

$$\begin{aligned}
&Maximize\ \ \tfrac{1}{\eta} \times \tfrac{1}{3}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4)\\
&+\tfrac{1}{1.52}\Big(-0.82 \times \textstyle\sum_{r=1}^{4}(v_1^{r1} - \frac{1}{4}\sum_{r=1}^{4} v_1^{r1}) \times \sum_{r=1}^{4}(v_1^{r3} - \frac{1}{4}\sum_{r=1}^{4} v_1^{r3})\\
&\qquad +0.33 \times \textstyle\sum_{r=1}^{4}(v_2^{r1} - \frac{1}{4}\sum_{r=1}^{4} v_2^{r1}) \times \sum_{r=1}^{4}(v_2^{r2} - \frac{1}{4}\sum_{r=1}^{4} v_2^{r2})\\
&\qquad -(1-\tfrac{0.6}{2}) \times \textstyle\sum_{r=1}^{4}(v_1^{r1} - \frac{1}{4}\sum_{r=1}^{4} v_1^{r1})^2\\
&\qquad -(1-\tfrac{1}{2}) \times \textstyle\sum_{r=1}^{4}(v_1^{r2} - \frac{1}{4}\sum_{r=1}^{4} v_1^{r2})^2\\
&\qquad -(1-\tfrac{0.4}{2}) \times \textstyle\sum_{r=1}^{4}(v_1^{r3} - \frac{1}{4}\sum_{r=1}^{4} v_1^{r3})^2\\
&\qquad -(1-\tfrac{0.5}{2}) \times \textstyle\sum_{r=1}^{4}(v_2^{r1} - \frac{1}{4}\sum_{r=1}^{4} v_2^{r1})^2\\
&\qquad -(1-\tfrac{0.5}{2}) \times \textstyle\sum_{r=1}^{4}(v_2^{r2} - \frac{1}{4}\sum_{r=1}^{4} v_2^{r2})^2\\
&\qquad -(1-\tfrac{1}{2}) \times \textstyle\sum_{r=1}^{4}(v_2^{r3} - \frac{1}{4}\sum_{r=1}^{4} v_2^{r3})^2\Big)\\
&s.t.\ \ E(\mathcal{U}), E_{d_r}^{\mathcal{A}_r},\\
&\qquad \varepsilon_r \geq 0.001, \forall r \in \{1,2,3,4\}.
\end{aligned} \tag{6}$$

Solving it for $\eta = 1$ leads to a simultaneous construction of value functions presented in Figure 4 for all decision makers. These functions exhibit some noticeable differences with respect to those obtained individually for each decision maker (see Figure 3). For example, slope of marginal value function over the first subinterval on X_1 is greater for d_4. In fact, the correlation between the first and third subintervals for X_1 is -0.82. Initially, the slope over the third subinterval for the value function of d_4 was less than the average slope, so we could expect a steeper slope compared to the group's average over the first subinterval. Nonetheless, the resulting slope over the first subinterval was equal to the mean.

The true comprehensive values of alternatives based on the assumed value functions as well as the values derived from the independent and joint estimations of preferences are presented in Table 5. The respective rankings are given in Table 6. Let us compare these rankings for each decision maker.

For d_1, the joint estimation could correctly separate a_1 from a_2, consistent with the true ranking, while the independent estimation failed to do so. Nonetheless, neither of the estimation procedures managed to separate a_2 and a_4. In fact, discriminating between these two alternatives is expected to be difficult when considering a small difference between their true comprehensive values (0.40 and 0.37). For d_2, the rankings derived from the two estimation procedures are

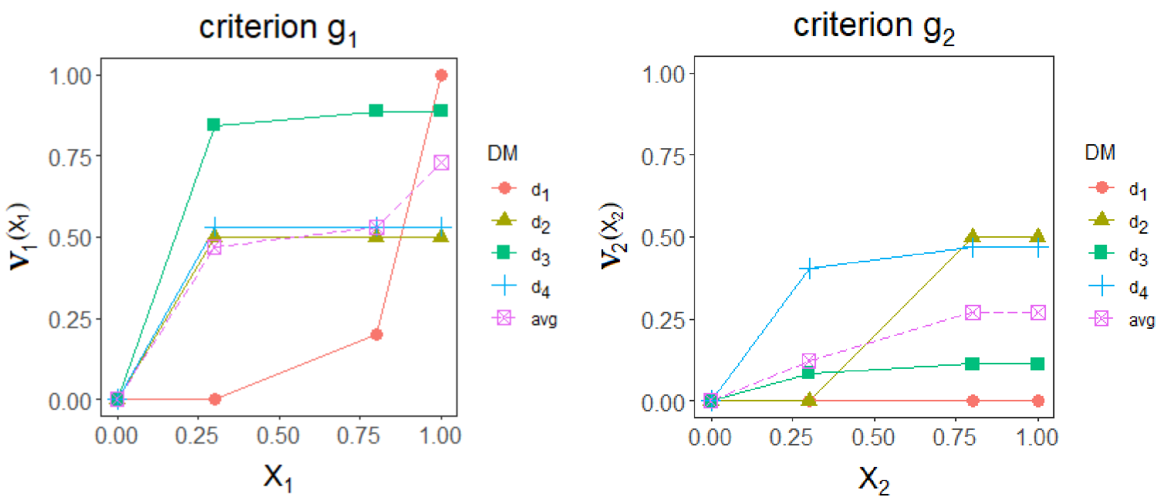

Figure 4: Value functions derived with a joint estimation of preferences by solving the convex optimization problem.

Table 5: Ground truth, independently estimated, and jointly estimated comprehensive values of alternatives.

	Ground truth				Independent estimation				Joint estimation			
	d_1	d_2	d_3	d_4	d_1	d_2	d_3	d_4	d_1	d_2	d_3	d_4
a_1	0.50	0.56	0.63	0.53	0.0	0.50	1.0	1.0	0.20	0.50	0.97	0.94
a_2	0.40	0.65	0.60	0.50	0.0	0.50	0.0	0.50	0.0	0.50	0.11	0.47
a_3	0.60	0.35	0.40	0.50	1.0	0.5	1.0	0.50	1.0	0.5	0.89	0.53
a_4	0.37	0.66	0.70	0.58	0.0	1.0	1.0	1.0	0.0	1.0	0.96	1.0

Table 6: Ground truth, independently estimated and jointly estimated rankings.

Decision Maker	Ground truth	Independent estimation	Joint estimation
d_1	$a_3 \succ a_1 \succ a_2 \succ a_4$	$a_3 \succ a_1 \sim a_2 \sim a_4$	$a_3 \succ a_1 \succ a_2 \sim a_4$
d_2	$a_4 \succ a_2 \succ a_1 \succ a_3$	$a_4 \succ a_2 \sim a_1 \sim a_3$	$a_4 \succ a_2 \sim a_1 \sim a_3$
d_3	$a_4 \succ a_1 \succ a_2 \succ a_3$	$a_1 \sim a_4 \sim a_3 \succ a_2$	$a_1 \succ a_4 \succ a_3 \succ a_2$
d_4	$a_4 \succ a_1 \succ a_2 \sim a_3$	$a_1 \sim a_4 \succ a_2 \sim a_3$	$a_4 \succ a_1 \succ a_3 \succ a_2$

identical. For d_3, in the joint estimation a_4 and a_3, as well as a_1 and a_3, are separated in the same way as in the true ranking, while the independent estimation failed to reproduce the preference relations for both these pairs. Finally, for d_4, when compared with the true preferences, the joint estimation correctly predicted the best choice and separated a_4 from a_1, whereas the independent estimation failed to do so.

When analyzing the rankings for all decision makers, the sole relation which is predicted better with the independent estimation is the indifference between a_2 and a_3 for d_4. For the values of Kendall's τ quantifying the similarities between the rankings derived from the independent and joint estimations with the true ranking, see Table 7. Specifically, the rankings for d_1 and d_3 are considerably improved when accounting for the structural patterns in the joint estimation of preferences. Note that these two are the decision makers with relatively less preference information compared to the group.

Table 7: Kendall's τ quantifying the similarities between the rankings obtained with the independent and joint estimations and the true ranking for each decision maker.

	$\tau_{independet}$	τ_{joint}
d_1	0.0	0.67
d_2	0.0	0.0
d_3	-0.33	0.33
d_4	0.67	0.67

4 Experimental Analysis Based on Simulation

In this section, we present a comprehensive experimental analysis based on a large number of randomly generated decision problems covering a variety of settings. The simulation parameters and their corresponding levels are presented in Table 8. In particular, we consider decision problems with different numbers of alternatives, criteria, decision makers, and pairwise comparisons provided by each decision maker. To simulate the holistic decision makers' preferences, two dimensions are considered: criteria weights and general shape of marginal value functions. Parameter λ defines the heterogeneity level among the decision makers in terms of criteria weights in their preferences – weights are interpreted as the maximal shares of criteria in the comprehensive value. In addition, the shapes of value functions are simulated as linear, concave or convex with different levels of curvature, or random. Finally, parameter η determines the intensity of incorporating structural patterns in the preference estimation procedure, i.e., a trade-off factor in the optimization problem (5).

Table 8: Simulation parameters and their corresponding levels used in the experimental analysis (R – number of decision makers, N – number of alternatives, M – number of criteria, P – number of pairwise comparisons provided by each decision maker, λ – heterogeneity level among decision makers, η – trade-off factor in the optimization problem).

Parameter	R	N	M	P	λ	Shape of value function	η
Levels	{5, 10, 15}	{6, 9, 12}	{3, 4, 5, 6}	{3, 5, 7}	{0.02, 0.1, 0.2, $+\infty$}	{linear, random, concave/convex}	{0.1, 0.4, 0.7, 1, 2, 4}

4.1 Simulation Design

The design of the simulation study consists of the following steps:

Step 1: Randomly generate N alternatives evaluated in terms of M criteria. The performances are randomly sampled from a uniform distribution in the range $[0, 1]$. The alternatives are verified to be distinct and incomparable in terms of a dominance relation.

Step 2: For each decision maker d_r, randomly select P distinct pairs of alternatives to be compared according to the decision maker's simulated preferences. We ensured that the relation for none of the pairs could be inferred from the transitivity of preferences or indifference relations. In case such a pair was detected, it was replaced by another randomly selected one.

Step 3: Preferences and rankings of alternatives for the decision makers are estimated independently and individually using the independent preference estimation approach, as well as by accounting for the structural patterns based on the proposed framework for the joint estimation of preferences. The two estimated rankings are compared against the decision maker's "true" ranking that is derived from the decision maker's simulated preference model. The results of this comparison refer to correctly predicting the top choice (best alternative) for the decision maker, indicated by α (equal to 1 if the top choice is predicted correctly) and Kendall's τ, accounting for the agreement level between all pairwise preference relations in the ranking. The reported values of α and τ are averaged over all decision makers.

To increase reliability and provide statistically invariant results, each decision setting is repeated 100 times, i.e., for each configuration, 100 different decision problems involving R decision makers are generated. In each replication, two values of Kendall's τ for each decision maker are computed, one for the rankings obtained with independent preference estimation (τ_{ind}), and the other for joint estimation of preferences (τ_{joint}). These values are averaged over the 100 replications and R decision makers. Analogously, we derive the proportion of decision scenarios for which the top choice is correctly predicted by either independent (α_{ind}) or joint (α_{joint}) estimation of preferences. Moreover, for each configuration, the differences between two measures are computed as $\delta\tau = \tau_{joint} - \tau_{ind}$ and $\delta\alpha = \alpha_{joint} - \alpha_{ind}$. A positive value of $\delta\tau$ ($\delta\alpha$) for a given configuration indicates an improvement gained in predicting a full ranking (a top choice) by employing the joint rather than the independent preference estimation approach.

4.2 Simulating Preference Models

The preference models are simulated along the following two dimensions: criteria weights and general shape of value functions.

Generating weights: For each decision maker, the criteria weights are generated considering different levels of heterogeneity among the decision makers. In the extreme case of a fully heterogeneous group of decision makers, these weights are generated from a uniform distribution independently for each decision maker seperately according to the following algorithm:

Step 1: Generate $M-1$ uniformly distributed random numbers $\theta_1, \theta_2, \cdots, \theta_{M-1}$ in the range $[0,1]$.

Step 2: Sort these numbers in an ascending order $\hat{\theta}_1 \leq \hat{\theta}_2 \leq \cdots \leq \hat{\theta}_{M-1}$. Also, define $\hat{\theta}_0 = 0$ and $\hat{\theta}_M = 1$.

Step 3: Compute weight of each criterion $g_m \in \mathcal{G}$ as $w^r_m = \hat{\theta}_m - \hat{\theta}_{m-1}$.

This extreme case in which decision makers form a fully heterogeneous group in terms of criteria weights is rather non-realistic. The criteria weights, e.g. relative importance of product attributes, might share a level of similarity across the decision makers and usually are not completely independent due to several factors including interaction among the decision makers, information diffusion, or context-dependent nature of some decision problems. In our simulation design, the level of heterogeneity among decision makers is directly controlled through an exogenous parameter that determines the dispersion of R points in an M-dimensional hyperspace, where each point corresponds to a weight vector for a decision maker, and M is the number of criteria.

To control the heterogeneity, we first generate a random weight vector in the M-dimensional hyperspace, which is treated as a reference point denoted by $\mathbf{w}^{ref}$. Then, for each $d_r \in \mathcal{D}$, we compute her criteria weights $\mathbf{w}^r = (w^r_1, \cdots, w^r_M)$ in the following way:

$$\mathbf{w}^r = \mathbf{w}^{ref} + \sum_{m=1}^{M} \delta_m \times (\mathbf{e}^m - \mathbf{w}^{ref}),\ r = 1, \cdots, R, \tag{7}$$

where $\mathbf{e}^m$ is a unit vector with its m^{th} component being equal to 1 and other components being 0, δ_m is the m^{th} component of a random vector $\mathbf{\Delta}$ that is drawn from a truncated multivariate normal distribution on the interval $[0, 1]$ with a mean $\mathbf{0}$ and a standard deviation $\lambda \mathbf{I}_M$, i.e., $\mathbf{\Delta} \sim \mathcal{N}_{[0,1]}(\mathbf{0}, \lambda \mathbf{I}_M)$, where $\mathbf{I}_M$ is an M-dimensional identity matrix.

We control the dispersion of R weight vectors, distributed around the reference point, by specifying the value of λ. This is demonstrated in Figure 5 for a decision problem with $M = 3$ criteria. The shaded triangle represents a space of all feasible weight vectors $[w_1, w_2, w_3]$ such that $w_1 + w_2 + w_3 = 1$ and $w_1, w_2, w_3 \geq 0$. The generated weights are conical combinations, with the coefficients δ_1, δ_2 and δ_3 of the three vectors connecting reference weight vector w^{ref} to the three vertices, hence always satisfying the normalization constraint $w_1 + w_2 + w_3 = 1$. In case condition $0 \leq w_m \leq 1$ is not satisfied for at least one criterion, the generation process is repeated. Coefficients $\delta_1, \delta_2,$ and δ_3 are drawn from $\mathcal{N}_{[0,1]}(\mathbf{0}, \lambda \mathbf{I}_3)$. Value of λ decides upon variation in magnitudes of conical combination coefficients and hence dispersion of weight vectors around the reference point $\mathbf{w}^{ref}$. This phenomenon is illustrated in Figure 6 for the case of two arbitrary levels of heterogeneity, which are applied to generate $R = 15$ weight vectors from the same reference point.

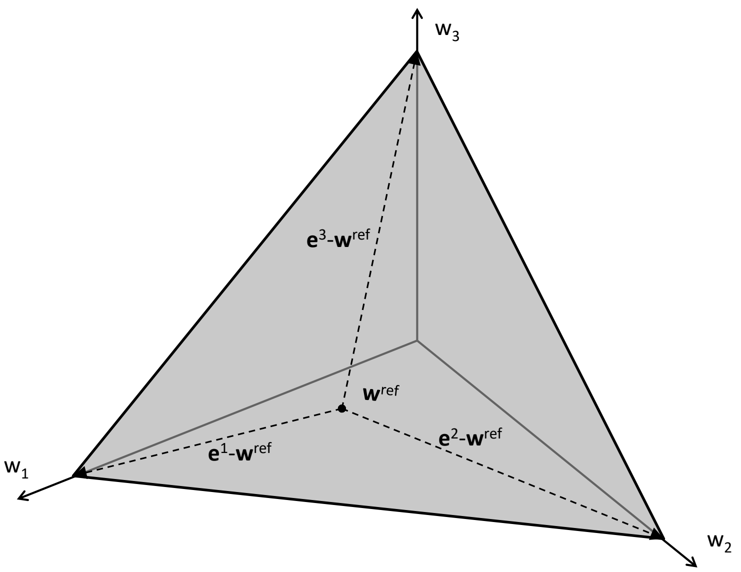

Figure 5: Generating weights by controlling heterogeneity for a problem involving $M = 3$ criteria.

Generating value functions: To ensure that there exists at least one value function compatible with the sampled P pairwise comparisons, alternatives are compared based on the simulated linear marginal value functions. To ensure that our results are not driven by a specific shape of value functions, we consider two additional general shapes for these functions: concave/convex and random. However, once the class of marginal value function shape is specified, all marginal value functions for all decision makers are simulated within the same class (hence being linear, or concave/convex, or completely random).

In case of concave/convex shape, each marginal value function is simulated as follows:

$$v(x) = \frac{1 - e^{-cx}}{1 - e^{-c}}, \tag{8}$$

where $c \neq 0$ is a Pratt-Arrow coefficient of absolute risk aversion and is drawn from a uniform

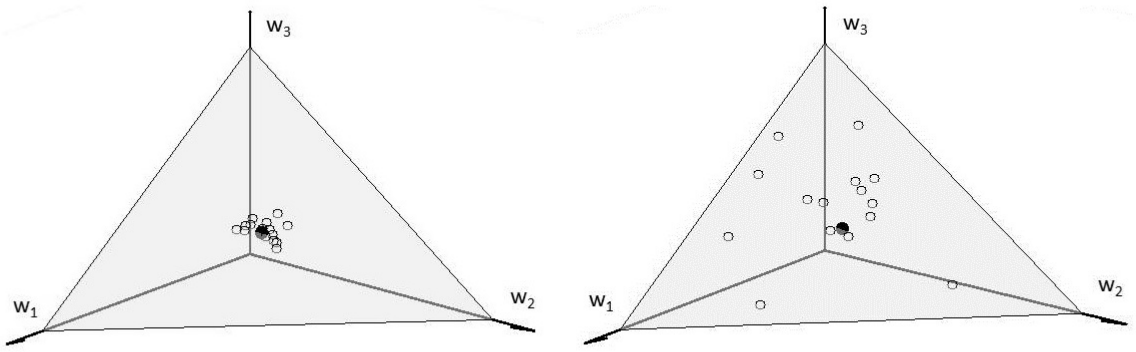

Figure 6: Weight vectors (empty circles) generated from the randomly selected reference weight vector (filled circle) for $R = 15$ decision makers and $M = 3$ criteria with $\lambda = 0.05$ (left) and $\lambda = 0.2$ (right) heterogeneity levels.

distribution over the interval $[-10, 10]$. This parameter determines curvature of the marginal value function. In fact, $c > 0$ ($c < 0$) corresponds to a concave (convex) function. For illustrative purpose, we present the marginal value functions for different levels of c in Figure 7.

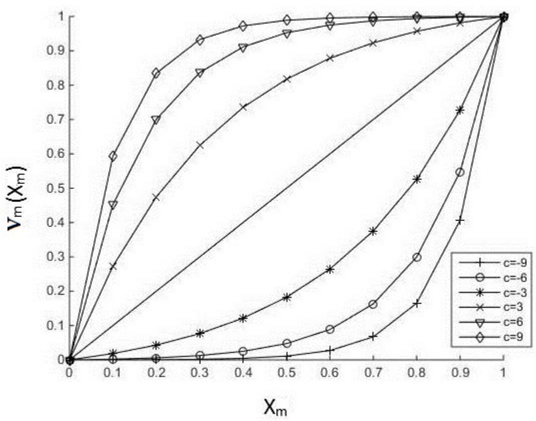

Figure 7: Simulated concave/convex marginal value functions for different curvature levels (a straight line in the middle corresponds to a risk neutral marginal value function $v(x) = x$).

In turn, a fully random shape of a marginal value function with $\alpha_m + 1$ breakpoints is simulated by generating $\alpha_m - 1$ random numbers in the interval $[0, 1]$, sorting them in an ascending order, $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{\alpha_m - 1}$, adding $\hat{\theta}_0 = 0$ and $\hat{\theta}_{\alpha_m} = 1$, and defining $v(x^j) = \hat{\theta}_j$, for $j = 0, 1, \cdots, \alpha_m$. The example random functions generated for $R = 5$ decision makers are presented in Figure 8.

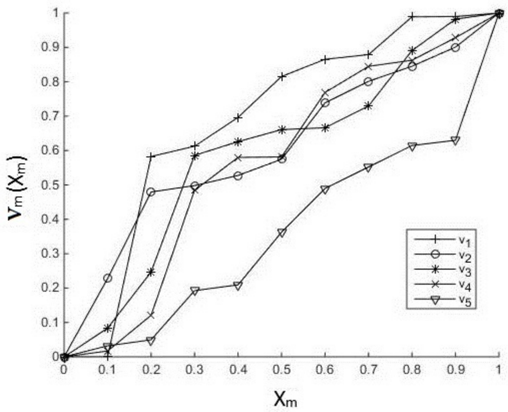

Figure 8: Simulated random marginal value functions for $R = 5$ decision makers.

After simulating criteria weights using either a completely heterogeneous approach or a specific level of λ, and marginal value functions with linear, concave/convex or random shapes for each $d_r \in \mathcal{D}$, the P randomly selected pairs of alternatives are compared in terms of their comprehensive values computed as follows: $V^r(a_n) = \sum_{m=1}^{M} w_m^r \cdot v_m^r(a_n)$. Overall, $1,296,000$ decision problems in $7,776$ different settings were generated.

4.3 Results

The distributions of improvements in prediction accuracies with respect to the full ranking ($\delta\tau$) and top choice ($\delta\alpha$) when employing a joint rather than an independent preference estimation method are presented in Figures 9 and 10, respectively. The respective descriptive statistics derived from the simulation study are provided in Table 9. They show a remarkable improvement in the prediction accuracy when accounting for the structural patterns in the estimation process. The improvement in α measure is by 0.022 (from 0.468 to 0.491), being statistically significant ($t = 92.7$ and $p - value < 0.001$), whereas the Kendall's τ measure is radically improved from 0.208 to 0.487 ($t = 226.6$ and $p - value < 0.001$).

Table 9: Descriptive statistics of the performance measures for the independent and joint estimations of preferences (count= 7776).

	Min	Max	Mean	S.E. mean	95% Conf. Interval	
τ_{ind}	−0.139	0.626	0.208	0.0023	0.203	0.213
τ_{joint}	0.096	0.777	0.487	0.0015	0.484	0.490
α_{ind}	0.224	0.792	0.468	0.0015	0.465	0.471
α_{joint}	0.226	0.812	0.491	0.0015	0.488	0.494
$\delta\tau$	0.049	0.589	0.279	0.0012	0.277	0.282
$\delta\alpha$	−0.117	0.126	0.022	0.0002	0.022	0.023

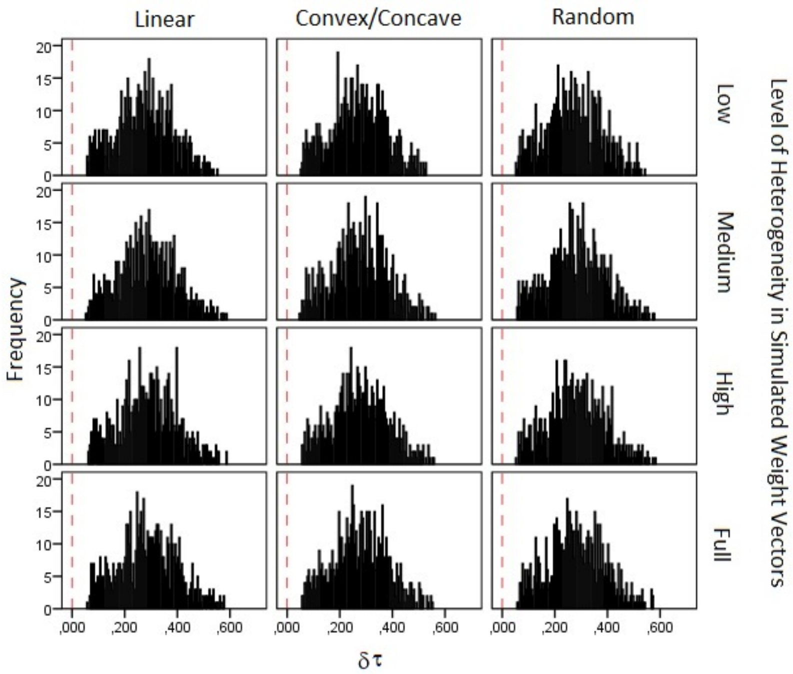

Figure 9: Distribution of gain in Kendall's τ when employing a joint preference estimation for different shapes of the simulated value functions and levels of heterogeneity in the simulated weight vectors (a dashed line is used to mark a zero value on the horizontal axis).

Figures 11 and 12 reveal that prediction accuracies in ranking and choice problems increase with the number of decision makers, and these results remain unchanged for different complexity

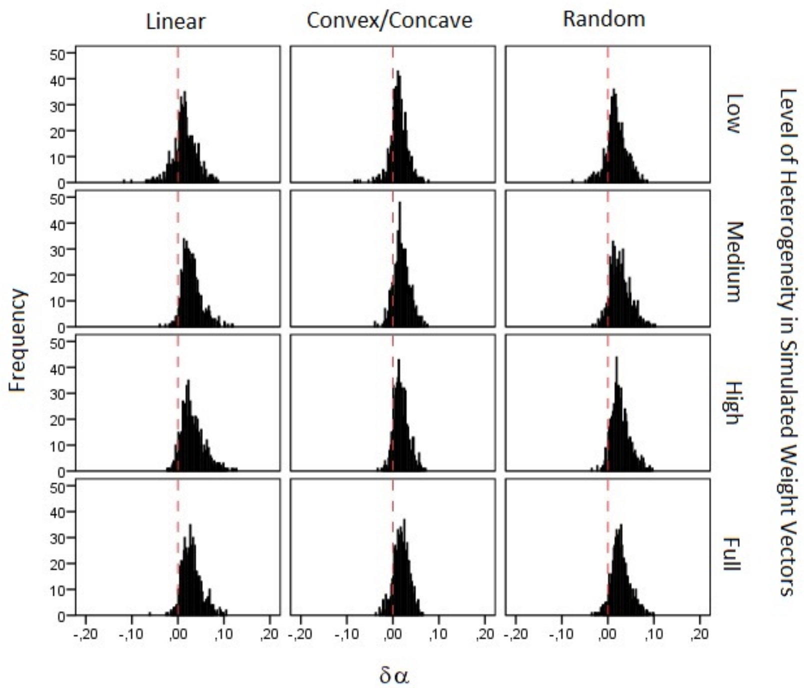

Figure 10: Distribution of gain in α measure when employing a joint preference estimation method for different shapes of the simulated value functions and levels of heterogeneity in the simulated weight vectors (a dashed line is used to mark a zero value on the horizontal axis).

levels in the simulated value functions. When a larger number of decision makers are involved in the decision making process, richer information on the structural patterns in value functions become available, resulting in better prediction accuracies in joint preference estimation method. On the contrary, the number of decision makers does not have any impact on the performance of the independent preference estimation approach.

Figures 13 and 14 reveal similar results for different levels of heterogeneity in the simulated weight vectors. The performance, in terms of prediction accuracy, of independent preference estimation method is not affected by the group size (R), whereas for the joint preference estimation approach the performance improves with a greater number of decision makers. Again, this conclusion is not affected by the level of heterogeneity in the simulated weight vectors.

Figures 15 and 16 show that the improvement in the predictive accuracy when employing

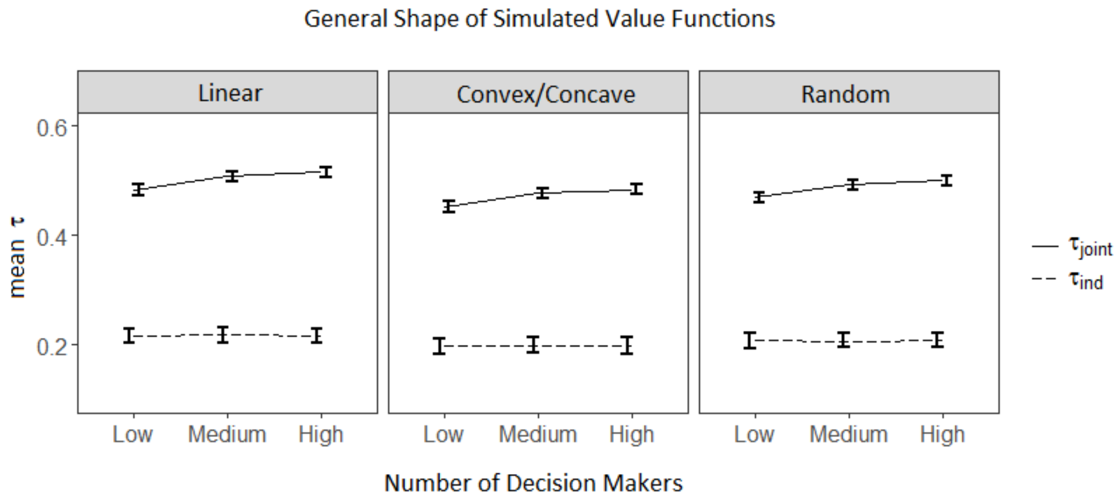

Figure 11: Average values of τ_{joint} (solid line) and τ_{ind} (dashed line) vs. the number of decision makers (R) for three general shapes of the simulated value functions (the bars indicate 95% confidence interval).

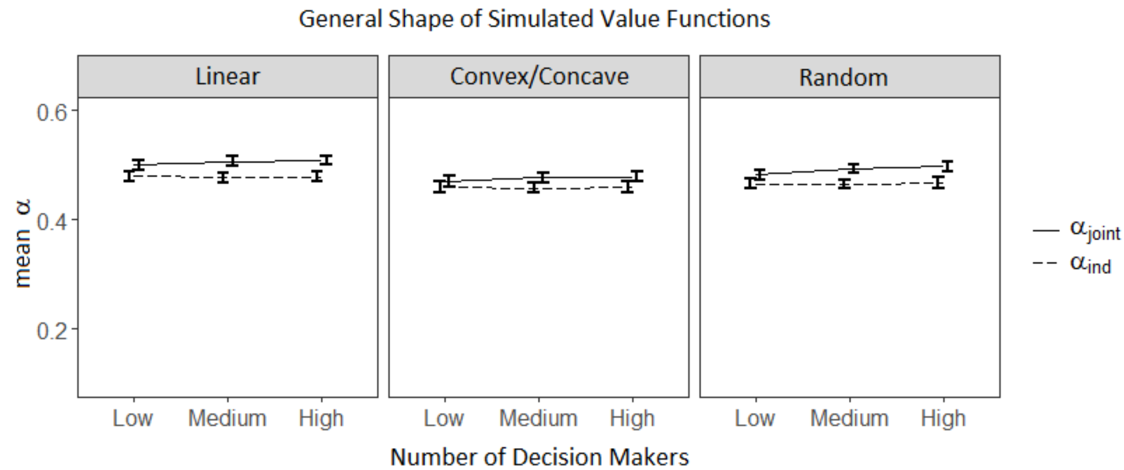

Figure 12: Average values of α_{joint} (solid line) and α_{ind} (dashed line) vs. the number of decision makers (R) for three general shapes of the simulated value functions (the bars indicate 95% confidence interval).

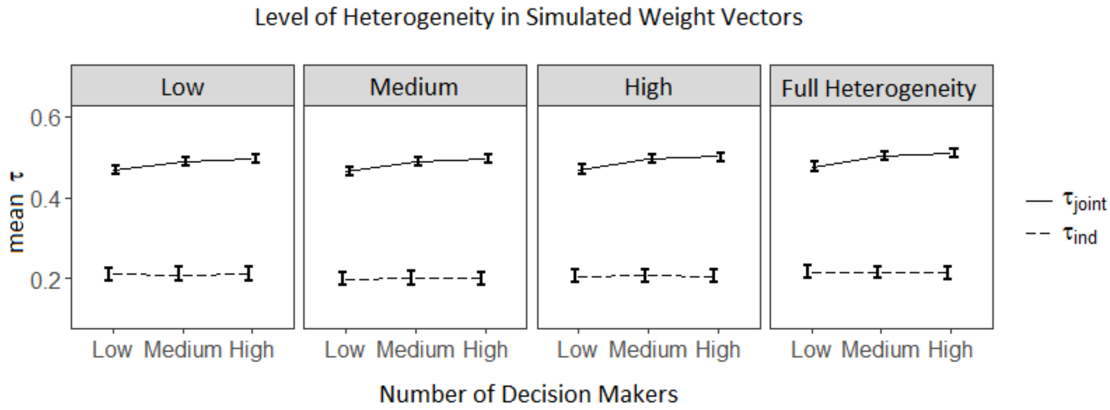

Figure 13: Average values of τ_{joint} (solid line) and τ_{ind} (dashed line) versus number of decision makers (R) for different levels of heterogeneity in simulated weight vectors (λ) (bars indicate 95% confidence interval).

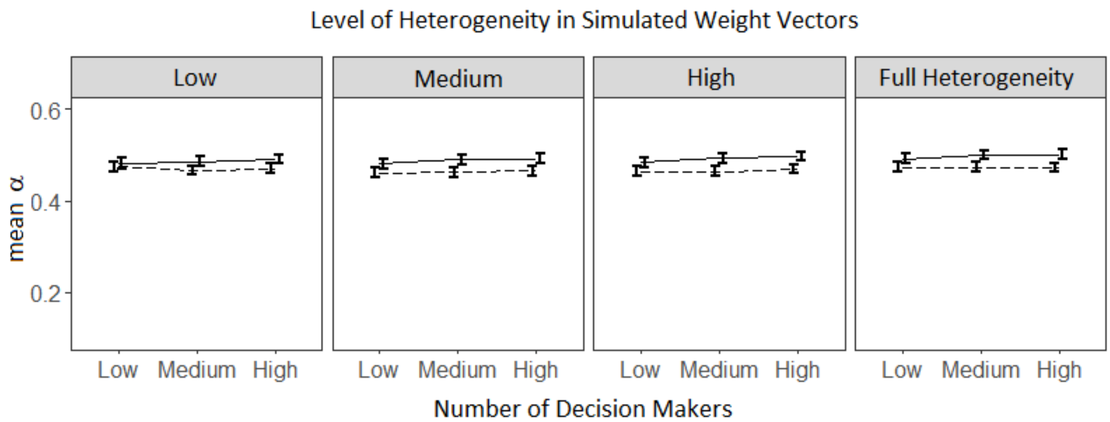

Figure 14: Average values of α_{joint} (solid line) and α_{ind} (dashed line) vs. the number of decision makers (R) for different levels of heterogeneity in simulated weight vectors (λ) (the bars indicate 95% confidence interval).

the joint preference estimation method ($\delta\tau$ and $\delta\alpha$) increases with a greater number of decision makers. Note that $\delta\tau$ is strictly increasing with R, whereas for most cases $\delta\alpha$ increases when passing from low to medium number of decision makers, while not changing much from medium to high number of decision makers. These findings are not affected by shape or heterogeneity factors. This observation demonstrates applicability of the proposed framework to a wide range of decision contexts irrespective of the composition of the group of decision makers and complexity levels in their value systems.

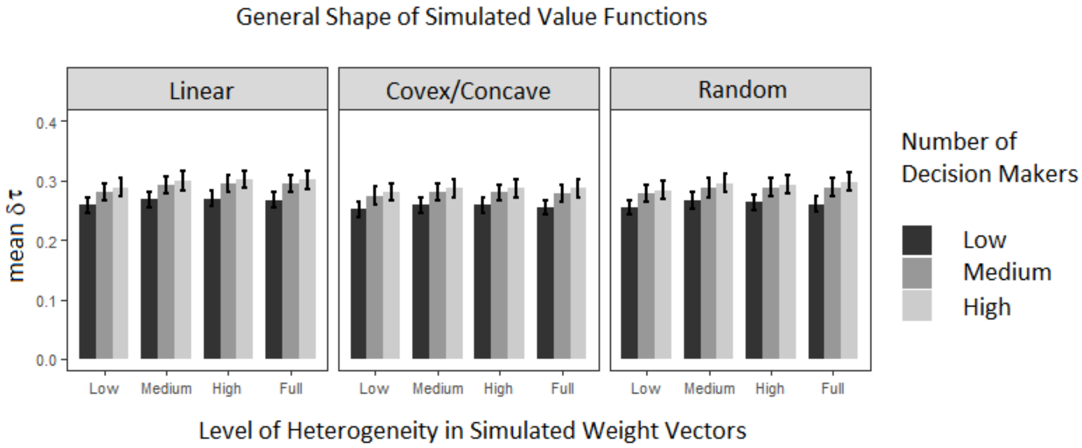

Figure 15: Average values of $\tau_{joint} - \tau_{ind}$ for different numbers of decision makers (blue – low; yellow – medium; purple – high) for different levels of heterogeneity in the simulated weight vectors (λ) and complexity levels in the simulated value functions (linear – left; concave/convex – middle; random – right) (the bars indicate 95% confidence interval).

The above discussed results demonstrate that employing a joint preference estimation method always improves predictive accuracy for the estimation of an entire ranking. The improvement in predicting a top choice, although being overall considerable and statistically significant, is not guaranteed for all scenarios as indicated, e.g., in Figure 10. In fact, for 9.8% of 7,776 decision settings (i.e., 765 cases) the α measure is deteriorated when applying a joint preference estimation method. Out of this, 507 cases (i.e., 6.6%) involve a small number of decision makers. Hence, we do not recommend the joint preference estimation approach for decision problems focusing solely on the top choice when the group's size is very small. Figure 17 shows that particularly in decision settings with a small number of decision makers, there exists an optimal level of parameter η that implies the highest improvement in the correct prediction of a top choice. Thus, even for such a disadvantageous setting, the proposed method can slightly improve the predictive accuracy in case the value of η is set properly (e.g., using a cross validation

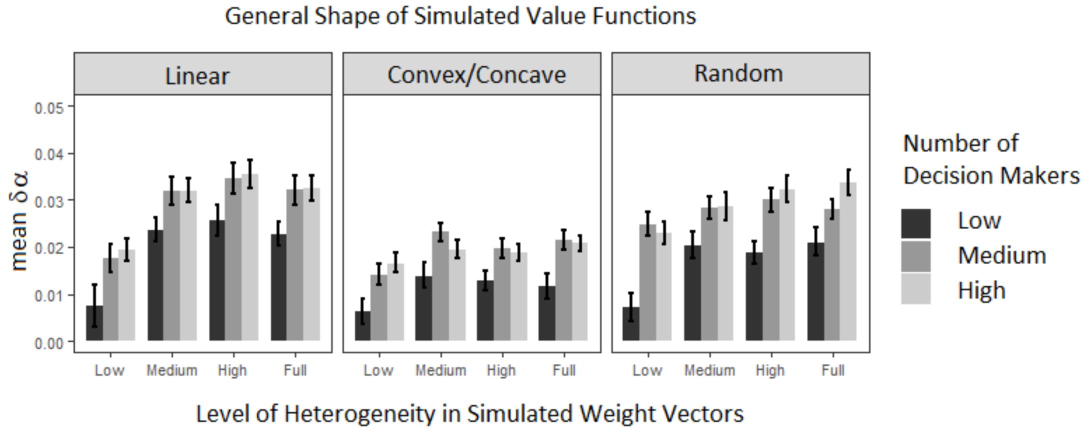

Figure 16: Average values of $\alpha_{joint} - \alpha_{ind}$ for different numbers of decision makers (blue – low; yellow – medium; purple – high) for different levels of heterogeneity in the simulated weight vectors (λ) and complexity levels in the simulated value functions (linear – left; concave/convex – middle; random – right) (the bars indicate 95% confidence interval).

technique).

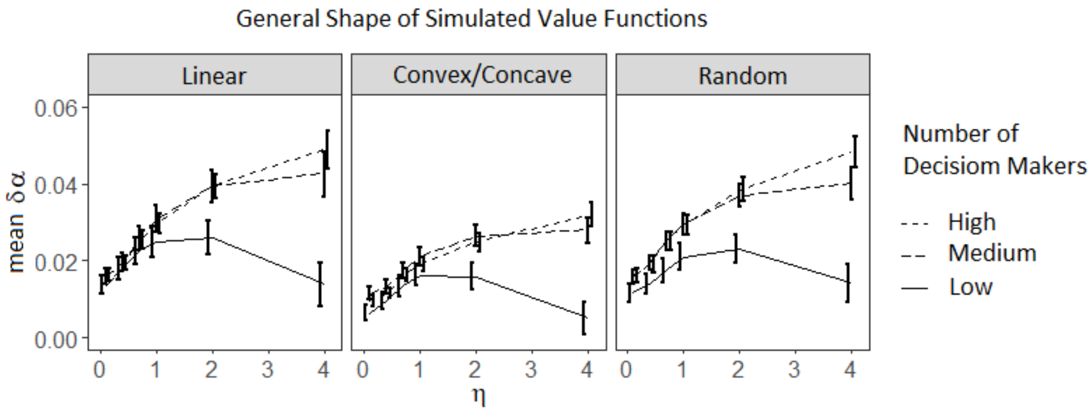

Figure 17: Average values of $\alpha_{joint} - \alpha_{ind}$ for different numbers of Decision Makers (solid – low; long dash – medium; dashed – high) and shapes of the simulated value functions (linear – left; convex/concave – middle; random – right) (the bars indicate 95% confidence interval).

Finally, we found that both measures of predictive validity (τ and α) for both joint and independent estimation methods decreased with an increase in the numbers of alternatives and criteria and a decrease in the number of supplied pairwise comparisons. Nevertheless, we observed that compared to an independent estimation method, the decline in τ_{joint} and α_{joint} is less when increasing the number of alternatives (i.e., $\delta\tau$ and $\delta\alpha$ are increasing with the number of alternatives). However, regarding the sensitivity of the two methods to the numbers of criteria and pairwise comparisons, we observed the opposite trend in τ and α measures. We found that τ_{joint} (α_{joint}) was less (more) affected compared to τ_{ind} (α_{ind}) with greater number of criteria (i.e., $\delta\tau$ ($\delta\alpha$) is increasing (decreasing) with the number of criteria). Conversely, α_{joint} increased more rapidly than α_{ind} when more pairwise comparisons were available, while the pattern for τ was inverse.

5 Analysis of Real Decision Makers' Preferences

In this section, we analyze the preferences of real decision makers using the independent and joint preference estimation methods, and compare the respective predictive accuracies. There are two crucial differences between the analysis of preferences elicited from real decision makers and the simulated preferences that are described in Section 4.2. First, data from real decision makers is typically noisy. Second, it does not provide any explicit information on the level of heterogeneity among the decision makers or complexity in their value systems.

We report the results of two studies based on the preferences from real decision makers. In the first study, we investigate the predictive accuracies of the independent and joint preference estimation methods while controlling the amount of preference information supplied by each decision maker. In the other study, we allow for the heterogeneity in the number of provided choice examples, as is typical for real-world problems, and compare the two methods across decision makers, while accounting for the idiosyncratic factors. In both studies, the impact of preference structure (i.e., choices of pairs of alternatives to be compared by each decision maker) is controlled through replications in sampling of pairs and averaging the results. Moreover, in the second study, we account for two sources of variations in predictive accuracy measures, the idiosyncratic factors related to the individual decision maker, and relative amount of preference information provided by each decision maker compared to other decision makers, by replicating across both dimensions. To account for sources of variation in predictive accuracy measures, in addition to sample design, we also employ hierarchical linear models (Luke, 2004; Hox et al., 2017) for the analysis of data in both studies.

5.1 Data

Preference information was collected from 94 real decision makers who were asked to rank 10 real phone contracts from the leading Polish mobile network operators. The packages were described based on four criteria (see Table 10). The preference elicitation was conducted following an active learning strategy selecting pairwise questions with the greatest potential information gain (Ciomek et al., 2017). The subjects were students from different programs at Poznan University of Technology. They answered the questions using a dedicated desktop application during independent sessions that were conducted in different time periods within an interval of one month. The preference elicitation was conducted until a ranking of phone contracts was obtained for each individual decision makers. In what follows, we call each participant of the experiment as a decision maker.

5.2 Study 1: the same amount of preference information provided by each decision maker

In the first study, we consider settings with the same number of pairwise comparisons provided by each decision maker. Specifically, we randomly select $P = 3$, $P = 5$ or $P = 7$ pairs of phone contracts compared in the same way as in the true ranking of $N = 10$ alternatives obtained for each decision maker. These sampled comparisons are treated as an input for the independent and joint preference estimation methods with different values of parameter η. For each decision maker, the ranking obtained with each method is compared with the decision maker's true ranking in terms of Kendall's τ and α measures. For each decision maker, we also derive values of ε obtained with the independent and joint estimated methods, denoted by ε^*_{ind} and ε^*_{joint},

Table 10: Performances of 10 real phone contracts in terms of four criteria.

Alternative	Fee per month g_1 (cost)	Contract's length g_2 (cost)	Internet data g_3 (gain)	Roaming calls in EU (min.) g_4 (gain)
a_1	30	20	1	0
a_2	70	12	5	400
a_3	50	12	5	120
a_4	40	12	1	0
a_5	130	24	20	400
a_6	50	24	7	0
a_7	60	20	5	400
a_8	70	12	7	120
a_9	60	24	10	0
a_{10}	80	12	10	120

respectively. To ensure that these measures are not affected by the choice of phone contracts involved in the input comparisons, the procedure is repeated 30 times with different pairs of alternatives. In this way, the differences in Kendall's τ and α among the decision makers reflect different levels of difficulty in reproducing their preferences using an additive value function, which, in turn, is implied by various levels of complexities in their value systems or heuristics that they employ in their judgments to evaluate the alternatives.

For each decision maker and each method, we report the values of ε, Kendall's τ and α averaged over 30 replications. For illustrative purpose, we provide the results for a single decision maker (d_1) in Table 11 for the 18 different configurations involving 3×6 levels of P and η, respectively. The complete table of results consists of measurements from 94 decision makers, thus, 1692 measurements in overall.

Model Specification and Analysis: The predictive accuracies of the independent and joint preference estimation methods are analyzed through graphical inspection and hierarchical linear models to control the individual differences. Multiple observations from the same decision maker cannot be regarded as independent from each other. Since the idiosyncratic factors and individual characteristics of decision makers (e.g., complexity level in their value system or

Table 11: Results of the experimental analysis in Study 1 (averaged over 30 replications for each level of P) for decision maker d_1 for different levels of P and η.

Decision Maker	P	η	ε_{ind}	τ_{ind}	α_{ind}	ε_{joint}	τ_{joint}	α_{joint}	$\delta\varepsilon$	$\delta\tau$	$\delta\alpha$
d_1	3	0.1	0.61	0.08	0.13	0.61	0.29	0.17	0.00	0.21	0.03
d_1	3	0.4	0.61	0.08	0.13	0.60	0.33	0.17	0.00	0.25	0.03
d_1	3	0.7	0.61	0.08	0.13	0.58	0.48	0.33	-0.03	0.40	0.20
d_1	3	1	0.61	0.08	0.13	0.54	0.55	0.33	-0.06	0.47	0.20
d_1	3	2	0.61	0.08	0.13	0.45	0.63	0.50	-0.15	0.55	0.37
d_1	3	4	0.61	0.08	0.13	0.35	0.69	0.57	-0.26	0.60	0.43
d_1	5	0.1	0.41	0.32	0.40	0.41	0.49	0.37	0.00	0.17	-0.03
d_1	5	0.4	0.41	0.32	0.40	0.41	0.52	0.40	0.00	0.19	0.00
d_1	5	0.7	0.41	0.32	0.40	0.40	0.57	0.37	-0.01	0.24	-0.03
d_1	5	1	0.41	0.32	0.40	0.39	0.63	0.43	-0.02	0.31	0.03
d_1	5	2	0.41	0.32	0.40	0.34	0.71	0.50	-0.07	0.39	0.10
d_1	5	4	0.41	0.32	0.40	0.24	0.73	0.60	-0.17	0.41	0.20
d_1	7	0.1	0.30	0.45	0.53	0.30	0.58	0.60	0.00	0.13	0.07
d_1	7	0.4	0.30	0.45	0.53	0.29	0.60	0.63	0.00	0.15	0.10
d_1	7	0.7	0.30	0.45	0.53	0.29	0.63	0.60	-0.01	0.19	0.07
d_1	7	1	0.30	0.45	0.53	0.28	0.66	0.60	-0.02	0.21	0.07
d_1	7	2	0.30	0.45	0.53	0.23	0.73	0.67	-0.07	0.28	0.13
d_1	7	4	0.30	0.45	0.53	0.17	0.78	0.80	-0.13	0.33	0.27

employed decision heuristic) might have an influence on effectiveness of the preference estimation methods, we employ hierarchical linear models to control for such factors.

In the employed multilevel modelling, the analysis is conducted at the following two levels: the level of individuals, I; level 1, and the level of decision setting, defined by (P, η); level 2. First, we examine Intraclass Correlation Coefficient (ICC) for all nine variables of interest (i.e., τ, α, and ε for the two methods, as well as $\delta\tau$, $\delta\alpha$, and $\delta\varepsilon$). If the ICC measure or observations from the graphical inspections indicate variation among the individuals, we employ a random intercept model with P as predictor in one model, and P and η as predictors in another model. We cannot treat the impact of η as a random effect because number of observations from each decision maker at each level of η is small (three) hence the estimated effect sizes will not be reliable (Snijders, 2005). The models are compared and validated using the likelihood ratio tests (changes in deviances) and based on the Schwarz's Bayesian Information Criterion (BIC) (Schwarz, 1978). For all validated models, we examined the residual plots and observed no obvious deviation from homoscedasticity or normality. For each variable of interest, we test the following two models in the same order of presentation:

$$
\begin{aligned}
M_1^y: \quad & y_{ij} = \beta_{0j} + \textstyle\sum_t \beta_1^{(t)} P^{(t)} + r_{ij}, && r_{ij} \sim N(0, \sigma_r^2) \\
& \beta_{0j} = \gamma_{00} + u_{0j}, && u_{0j} \sim N(0, \sigma_{u_0}^2) \\
\\
M_2^y: \quad & y_{ij} = \beta_{0j} + \textstyle\sum_t \beta_1^{(t)} P^{(t)} + \sum_q \beta_2^{(q)} \eta^{(q)} + r_{ij}, && r_{ij} \sim N(0, \sigma_r^2) \\
& \beta_{0j} = \gamma_{00} + u_{0j}, && u_{0j} \sim N(0, \sigma_{u_0}^2)
\end{aligned}
\tag{9}
$$

$$y \in \{\tau_{ind}, \alpha_{ind}, \varepsilon_{ind}, \tau_{joint}, \alpha_{joint}, \varepsilon_{joint}, \delta\tau, \delta\alpha, \delta\varepsilon\},$$

where y_{ij} is the observation for configuration $i \in \eta \times P$ for individual j. Variables P and η are treated as categorical with the least values as baseline, and dichotomous variables $P^{(t)}$ and $\eta^{(q)}$ are defined to represent their other values, where $P^{(t)} = 1$ if $P = t$, and 0 otherwise; similarly $\eta^{(q)} = 1$ if $\eta = q$, and 0 otherwise. Note that the models M_2^y are tested only for the observations from the joint preference estimation method, as measurements from the independent preference estimation method are not influenced by the parameter η.

Results: The descriptive statistics derived from the analysis are presented in Table 12. The results show that the predictive validity measures for both ranking and choice problems are improved by employing the joint preference estimation method. In addition, a gain in the predictive validity is attained at the cost of model's expressiveness, which is quantified with ε.

The results show that particularly in predicting a full ranking, the gain in the predictive validity is remarkably large ($\bar{\delta\tau} = 0.299$, $t = 97.1$, $p - value < 0.001$), and it is relatively greater when smaller amount of preference information is available ($F = 235.8$, $p - value < 0.001$) as shown in Figures 18 and 19. Note that the gain in predictive validity is non-decreasing with η,

Table 12: Descriptive statistics for the analysis of real decision makers' preferences using the joint and independent preference estimation methods in Study 1 (count= 1692).

Measure	Min	Max	Mean	S.E. Mean	95% Conf. Interval	
τ_{ind}	-0.243	0.507	0.193	0.0041	0.185	0.201
τ_{joint}	-0.025	0.828	0.491	0.0039	0.484	0.499
α_{ind}	0.000	0.967	0.328	0.0061	0.315	0.340
α_{joint}	0.033	0.967	0.384	0.0040	0.376	0.392
ε_{ind}	0.205	0.728	0.429	0.0032	0.422	0.435
ε_{joint}	0.108	0.728	0.377	0.0032	0.370	0.383
$\delta\tau$	0.061	0.713	0.299	0.0031	0.293	0.305
$\delta\alpha$	-0.667	0.700	0.056	0.0044	0.048	0.065
$\delta\varepsilon$	-0.337	0.000	-0.052	0.0017	-0.055	-0.049

which is understandable in view of the simulation results from Section 4.2 when the number of decision makers is large ($R = 94$). We also observed that τ_{joint} obtained for $\eta \geq 1$ and only $P = 3$ pairwise comparisons given by each decision maker was greater than τ_{ind} for the case with $P = 7$ pairwise comparisons provided by each decision maker, i.e. $\tau_{joint}(P = 3, \eta \geq 1) \geq 0.403$ and $\tau_{ind}(P = 7) = 0.359$, as shown in Figure 20.

Graphical analysis and ICC values demonstrate that for τ_{ind}, τ_{joint}, α_{ind}, α_{joint}, and $\delta\alpha$, variations among the individuals are considerably large relative to the total variation. However, small ICC values are observed for ε_{ind}, ε_{joint}, $\delta\varepsilon$, and $\delta\tau$ (e.g., $ICC = 0.03$ for ε_{ind} and $ICC <$ 0.01 for others). This observation is particularly important when considering $\delta\tau$. Small variation of $\delta\tau$ across the individuals implies that even though accuracies in predicting a full ranking in both preference estimation methods differ from one decision maker to another and from one (η, P) configuration to another, the improvement in predictive validity when accounting for structural patterns does not vary across the individual decision makers. Such a finding supports generalizability of the joint preference estimation method in ranking problems, beyond decision makers' individual characteristics. This result is consistent with the findings from the simulation study, where the gain in the predictive validity was found to be independent from the level of complexity in the simulated value functions or heterogeneity in the weight vectors.

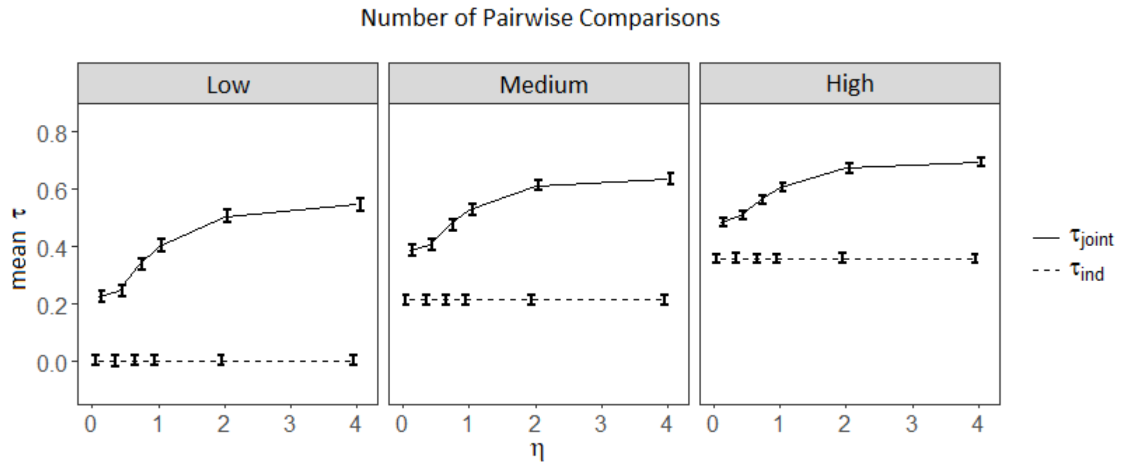

Figure 18: Average values of τ_{joint} (solid line) and τ_{ind} (dashed line) versus η for different numbers of pairwise comparisons provided by each decision maker in Study 1 (the bars indicate 95% Confidence Interval).

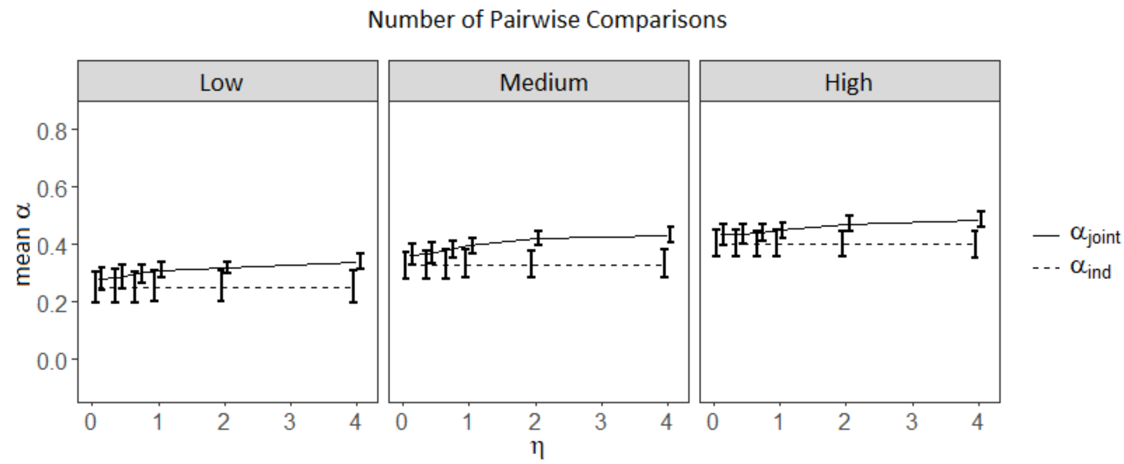

Figure 19: Average values of α_{joint} (solid line) and α_{ind} (dashed line) vs. the levels of η for numbers of pairwise comparisons provided by each decision maker in Study 1 (the bars indicate 95% Confidence Interval).

When comparing $M_1^{\tau_{joint}}$ and $M_2^{\tau_{joint}}$, we observed that η has a significantly positive impact on τ_{joint} ($\chi^2 = 2904, p-value < 0.001$) with a logarithmic pattern (i.e., the marginal influence

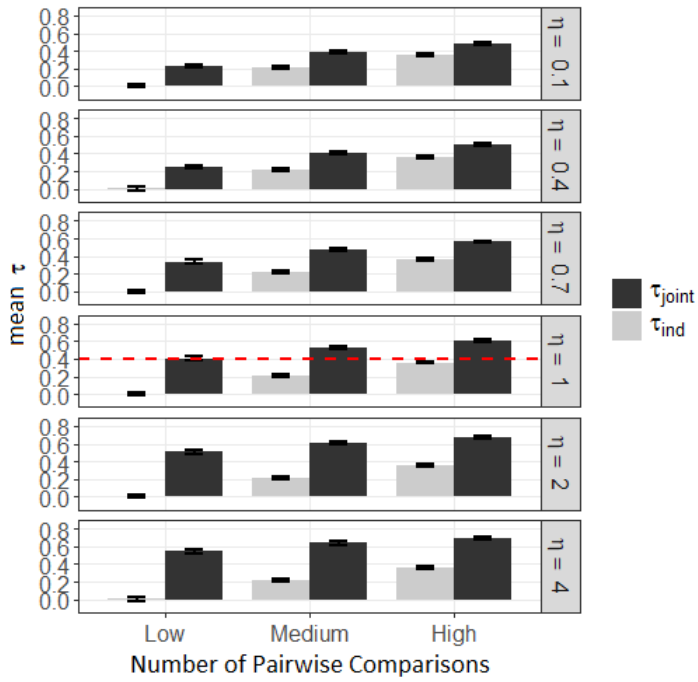

Figure 20: Comparison of τ_{joint} (black) and τ_{ind} (grey) for different levels of P and η in study 1 (the red dashed line compares τ_{joint} for $\eta \geq 1$ and $P = 3$ with τ_{ind} for $P = 7$).

is smaller when η grows). An in-depth analysis of τ_{joint} in a random intercept model with P and η as the categorical predictors (model $M_2^{\tau_{joint}}$) indicates a considerable individual variation in the intercepts (ranging from 0.02 to 0.39 with mean equal to 0.25 and St.D – 0.08), which reflects a high level of heterogeneity among the decision makers (St.D. of intercepts was observed to be 0.075 in $M_1^{\tau_{ind}}$). Furthermore, by allowing an individual variation for the slope coefficients corresponding to levels of η, no improvement was observed in the model fit (in fact, BIC increased from -5279 in model $M_2^{\tau_{joint}}$ to -5212 in a model by random intercepts and random slopes). This implies that the impact of η on predictive accuracy in estimating an entire ranking is independent from the individual characteristics of decision makers (see Figure 21 in Appendix A), even though – as suggested by the results of simulation study – it might still depend on the particular decision settings (e.g., the number of decision makers).

In contrast to $\delta\tau$, our results show that variation in $\delta\alpha$ is considerably large across the individuals ($ICC = 0.43$). However, $\delta\tau$ appeared to be significantly and negatively affected by

P, while no significant differences in $\delta\alpha$ were observed among different levels of P ($F = 1.03$, $p - value = 0.36$). This means that when more preference information becomes available, τ_{ind} grows faster than τ_{joint}, but the impacts on α_{ind} and α_{joint} are similar (see Figures 18 and 19). Accounting for these two observations, we test model $M_2^{\delta\alpha}$ and by excluding P. As a result, BIC decreased from -2526 to -2534. Even though the impact of η on $\delta\alpha$ is positive at the aggregate level, a graphical analysis reveals that this impact differs across the individuals in both magnitude and direction (see Figure 22 in Appendix A). That is, the use of greater values of η results in larger values of $\delta\alpha$ for most individuals, but for some others – this impact is negative or appears to follow a trend of an inverted "U". In support of this observation, BIC decreased considerably by allowing random slopes in $M_2^{\delta\alpha}$ (from -2534 to -3360), demonstrating a variation in the impact of η on $\delta\alpha$ among the individuals. However, the estimated effect sizes at the individual level for this model are not reliable due to a small number (three) of observations for each decision maker at each level of η (Snijders, 2005). Therefore, we cannot draw any concrete conclusions from this model. This issue will be addressed in the other part of the study in Section 5.3.

Results of the multilevel analysis are presented in Table 13. They demonstrate a remarkably large improvement in predictive accuracy for ranking problem. The intercept for τ_{joint} is considerably larger than that of τ_{ind}, with approximately the same level of uncertainty attributed to superpopulation, i.e. the estimates of σ_{u_0}. Moreover, η demonstrates a significantly positive impact on τ_{joint} with a large effect size. Note that based on the estimated coefficients in the multilevel models, the predicted value for τ_{joint} with $P = 3$ and $\eta = 1$ is larger than that of τ_{ind} with $P = 7$ ($0.253 + 0.146$ versus 0.355), which confirms our earlier observation from graphical analysis (see Figure 20).

When considering $\delta\alpha$, the improvement is significant only when η is sufficiently large. This is consistent with our finding from the simulation analysis that the improvement in predictive accuracy for choice problem is increasing with η in case of a sufficiently large number of decision makers being involved (see Figure 17). Despite the positive and significant impact of η on $\delta\alpha$ at the group level, large variation in the intercepts at the individual level ($\sigma_{u_0} = 0.149$) as well as relatively small effect sizes for levels of η (ranging from 0.005 for $\eta = 0.4$ to 0.061 for $\eta = 4$

Table 13: Maximum likelihood estimates of parameters of multilevel models in Study 1, and OLS estimates for $\delta\tau$ (tests of significance in multilevel models are performed by Satterthwaite's method).

		τ_{ind}	τ_{joint}	$\delta\tau$	α_{ind}	α_{joint}	$\delta\alpha$
Fixed-Effects							
	Intercept	.004	.253***	.249***	.253***	.280***	.031
	$P^{(5)}$	.211***	.131***	−.080***	.076***	.088***	
	$P^{(7)}$	.355***	.211***	−.145***	.149***	.147***	
	$\eta^{(0.4)}$		.020***	.020***		.005	.005
	$\eta^{(0.7)}$		.091***	.091***		.015	.015
	$\eta^{(1)}$		.146***	.146***		.026**	.026**
	$\eta^{(2)}$		.231***	.231***		.045***	.045***
	$\eta^{(4)}$		.259***	.259***		.061***	.061***
Random Effects							
	σ_r	.037	.044		.073	0.106	.102
	σ_{u_0}	.075	.082		.234	0.111	.149
Model Fit							
	BIC	−5878	−5279		−3542	−2425	−2534
	$Log-Likelihood$	2958	2677		1789	1249	1297

Significance codes: < 0.001 – ***, 0.001 – ** , 0.01 – *

For $\delta\tau$, adjusted $R^2 = 0.816$

in a logarithmic pattern) imply high level of uncertainty in the overall value of $\delta\alpha$.

Finally, similar to τ_{joint}, we observe that α_{joint} increases – with the same pace as α_{ind} – when more preference information is available. However, this impact is smaller in magnitude compared to the case of τ_{joint}. In other words, after accounting for the individual differences, more preference information has a larger positive impact on the predictive accuracy for ranking problems (τ_{ind} and τ_{joint} measures) than for choice problems (α_{ind} and α_{joint} measures).

5.3 Study 2: different amount of preference information provided by each decision maker

In this section, we use the same data on preferences of real decision makers to examine the predictive accuracies of the two methods by allowing variation among the decision makers in terms of the amount of provided input information. This is closer to a typical real world situation where the amount of supplied preferences differs from one individual to another. We are interested in verifying how contributions of each individual in the supplied preference information (i.e., relative amount of preference information provided by the individual compared to the group) affects the predictive accuracy of the constructed value function. Specifically, we aim at answering how such an accuracy is affected if, relative to the group, there is less information available about the decision maker's preferences compared to other individuals; also how well the joint preference estimation method performs in capturing heterogeneity in value functions in such settings.

To simulate a scenario with unbalanced preference information among decision makers, we first randomly assigned a number of pairwise comparisons $p_r \in \{3, 5, 7\}$ to be sampled from the ranking of $N = 10$ phone contracts for each of $R = 94$ individual decision makers $d_r \in \mathcal{D}$. Then, the individual value functions were constructed using the independent and joint preference estimation methods, and the respective predictive accuracies were quantified in terms of Kendall's τ and α measures. To ensure that these measures were not affected by the choice of phone contracts for pairwise comparisons, the results were averaged over 30 runs with different randomly selected p_r pairs for each decision maker. The analysis was repeated 100 times with different values of p_r assigned to the decision makers. Note than p_r is fundamentally different than parameter P used in Study 1, even though they both represent the numbers of provided pairwise comparisons. The parameter p_r demonstrates a relative amount of preference information provided by a particular decision maker compared to the group, while P represents a total amount of preference information available from the group that is balanced over individuals.

Model Specification and Analysis: In the joint estimation of preferences with heterogeneous amount of preference information from the decision makers, one concern is that the constructed value functions are biased towards the decision makers who have provided richer preference information. In such case, the constructed value functions for the decision makers with relatively less amount of preference information, might be dominated by those with more preference information, hence the heterogeneity in preferences would not be captured adequately in the estimated preference models. To address this concern, we compare improvements in predictive validity of constructed value functions when employing joint preference estimation method between groups of decision makers with different amount of supplied preference information. For this purpose, in our analysis, we controlled two sources of variation in predictive accuracy measures: variation due to the individual differences between decision makers, and variation due to the relative position of a decision maker compared to the group in terms of number of supplied pairwise comparisons. A graphical inspection presented in Section B shows these two sources of variation are considerably large for most measures.

The graphical analysis confirms a high variation by individual for all the predictive measures but $\delta\tau$. In fact, the analysis of variance demonstrates that variation by individual accounts only for 6% of the total variation in $\delta\tau$ – whereas in case of P it is already 16%. This observation supports our finding from the simulation study and from the analysis of preferences in Section 5.2 related to the gain in the predictive accuracy for ranking problems being independent from the individual characteristics of decision makers.

Conversely, the improvement in the predictive accuracy measures for choice problems, i.e. $\delta\alpha$, demonstrates a high variation by individuals and a small variation by P. In other words, such a gain appears to be independent from the decision maker's relative position with respect to the amount of supplied preference information. Note that with more pairwise comparisons provided by the decision maker, the predictive accuracy measures τ_{ind}, τ_{joint}, α_{ind}, and α_{joint} are greater. However, the magnitude of impact is comparable for both α_{ind} and α_{joint}, resulting in a similar gain in the predictive accuracy for different decision makers, irrespective of whether they provide more or less preference information compared to the group.

Further inspection reveals that the impact of parameter η on $\delta\tau$ and τ_{joint} is similar across both individuals and levels of P, as presented in Figures 33 and 34, respectively. Both measures increase with η following a logarithmic trend for all individuals and with P (parallel curves for different levels of P). Taking into account the variation across individuals and across relative preference information groups, as well as an invariance in effect of η over the individuals and relative preference information groups, the following multilevel models are specified to examine the variation in $\delta\tau$ and τ:

$$\begin{aligned}
&\delta\tau_{ip} = \beta_{0p} + \textstyle\sum_q \beta_1^{(q)}\eta^{(q)} + r_{ip}, && r_{ip} \sim N(0, \sigma_r^2) \\
&\beta_{0p} = \gamma_{00} + v_{0p}, && v_{0p} \sim N(0, \sigma_{v_0}^2) \\
&\tau_{joint_{ijp}} = \beta_{0jp} + \textstyle\sum_q \beta_1^{(q)}\eta^{(q)} + r_{ijp}, && r_{ijp} \sim N(0, \sigma_r^2) \\
&\beta_{0jp} = \gamma_{000} + u_{0j0} + v_{00p}, && u_{0j0} \sim N(0, \sigma_{u_0}^2), v_{00p} \sim N(0, \sigma_{v_0}^2)
\end{aligned} \tag{10}$$

where $\delta\tau_{ip}$ is the i^{th} observation at level p of relative preference information, and τ_{ijp} is the i^{th} observation for individual j at level p of relative preference information. Terms v_{0p} and v_{00p} capture variation in $\delta\tau$ and τ_{joint}, respectively, by relative supplied preference information, whereas term u_{0j0} captures variation by individual in τ_{joint}.

In contrast to the τ measures, the impact of η on $\delta\alpha$ and α_{joint} was observed to vary by individuals, as shown in Figures 35 and 36. Furthermore, even though a variation in α_{joint} is considerable between the groups defined by different levels of P – see Figure 32 – no strong evidence indicating a variation in the impact of η on α_{joint} by levels of P was observed – the curves plotting α_{joint} versus η in Figure 36 are approximately parallel for almost all individuals. Moreover, variation in $\delta\alpha$ was observed to be negligible by P, as discussed before. Taking into account these observations, the two following multilevel models are employed to investigate $\delta\alpha$ and α_{joint}, respectively:

$$\begin{aligned}
&\delta\alpha_{ij} = \beta_{0j} + \textstyle\sum_q \beta_{1j}^{(q)}\eta^{(q)} + r_{ij}, && r_{ij} \sim N(0, \sigma_r^2) \\
&\beta_{0j} = \gamma_{00} + u_{0j}, && u_{0j} \sim N(0, \sigma_{u_0}^2) \\
&\beta_{1j}^{(q)} = \gamma_{10}^{(q)} + u_{1j}, && u_{1j} \sim N(0, \sigma_{u_1}^2) \\
&\alpha_{joint_{ijp}} = \beta_{0jp} + \textstyle\sum_q \beta_{1j}^{(q)}\eta^{(q)} + r_{ijp}, && r_{ijp} \sim N(0, \sigma_r^2) \\
&\beta_{0jp} = \gamma_{000} + u_{0j0} + v_{00p}, && u_{0j0} \sim N(0, \sigma_{u_0}^2), v_{00p} \sim N(0, \sigma_{v_0}^2) \\
&\beta_{1j}^{(q)} = \gamma_{10}^{(q)} + u_{1j}^{(q)}, && u_{1j}^{(q)} \sim N(0, \sigma_{u_1^{(q)}}^2)
\end{aligned} \tag{11}$$

Note that the random intercepts vary by groups distinguished by different levels of preference information for $\delta\tau$, by individuals for $\delta\alpha$, by individuals and by groups defined by different levels of P for τ_{joint} and α_{joint}. Moreover, random effects of η vary by individuals for $\delta\alpha$ and α_{joint} measures. To estimate the fixed and random coefficients, we used a maximum likelihood estimation. Executions were performed using *lme4* library in the R package (Bates et al., 2014).

Results: The descriptive statistics derived from the analysis are presented in Table 14. The results show that, similarly to the scenarios with balanced preference information for different decision makers, the predictive validity for both ranking and choice problems is improved when employing the joint preference estimation method in heterogeneous preference information setting. Compared to the scenario with the same amount of pairwise comparisons for all decision makers, the predictive accuracy measures demonstrate wider range, but narrower confidence intervals of the mean at 95% level of confidence. However, no considerable difference in the mean values was observed.

Table 14: Descriptive statistics for the analysis of real decision makers' preferences using the joint and independent preference estimation methods with heterogenous amount of preference information in Study 2 (count= 56400).

Measure	Min	Max	Mean	S.E. Mean	95% Conf. Interval	
τ_{ind}	-0.271	0.550	0.194	0.0007	0.193	0.196
τ_{joint}	-0.119	0.861	0.489	0.0007	0.487	0.490
α_{ind}	0.000	1.000	0.327	0.0011	0.325	0.329
α_{joint}	0.000	1.000	0.380	0.0007	0.378	0.381
ε_{ind}	0.180	0.764	0.427	0.0006	0.426	0.428
ε_{joint}	0.086	0.764	0.375	0.0006	0.374	0.376
$\delta\tau$	0.027	0.787	0.294	0.0005	0.293	0.295
$\delta\alpha$	-0.767	0.767	0.052	0.0008	0.051	0.054
$\delta\varepsilon$	-0.376	0.000	-0.052	0.0003	-0.053	-0.051

Results from the analysis of multilevel models in (10) and (11) are presented in Table 15. By comparing the intercepts, while accounting for the variation by individuals and by relative

Table 15: Maximum likelihood estimation of parameters of multilevel models in Study 2 (tests of significance are performed by Satterthwaite's method).

		τ_{ind}	τ_{joint}	$\delta\tau$	α_{ind}	α_{joint}	$\delta\alpha$
Fixed-Effects							
	Intercept	.193	.364**	0.171*	.327***	.356***	.029***
	$\eta^{(0.4)}$		.020***	.020***		.005**	.005***
	$\eta^{(0.7)}$		.088***	.088***		.016**	.016***
	$\eta^{(1)}$		.145***	.145***		.023	.023***
	$\eta^{(2)}$		.230***	.230***		.040	.040***
	$\eta^{(4)}$		.261***	.261***		.058	.058***
Random Effects							
	σ_r	.043	.048	.057	.085	.089	.083
	σ_{u_0}	.072	.079		.229	.233	.047
	σ_{v_0}	.142	.092	.051	.061	.128	
	$\sigma_{u_1^{(0.4)}}$					.006	.006
	$\sigma_{u_1^{(0.7)}}$					.049	.026
	$\sigma_{u_1^{(1)}}$					.099	.054
	$\sigma_{u_1^{(2)}}$					.194	.096
	$\sigma_{u_1^{(4)}}$					.277	.139
Predicted v_{00p}							
	$P = 3$	-0,184	-0,120	.064	-0,074	-0,072	
	$P = 5$	.022	.019	-0.003	.005	.006	
	$P = 7$	.162	.101	-0.061	.069	.066	
Model Fit							
	BIC	-194649	-180815	-162196	-116802	-111150	-119005
	$Log-Likelihood$	97347	90457	81142	58423	55734	59656

Significance codes: < 0.001 – ***, 0.001 – ** , 0.01 – *

preference information groups, these results show that the improvement in predictive accuracy for ranking problem is remarkably large, and for choice problems it is significantly positive, though much smaller in magnitude.

The predicted values for the random coefficients related to the groups defined by the relative preference information, i.e. random parameter v_{00p}, show that the predictive accuracy for individuals with larger number of supplied pairwise comparisons, relative to the group, is higher than for those with relatively smaller number of pairwise comparisons, for both independent and joint preference estimation methods. Moreover, the effect sizes are roughly similar for the α measures, and are not very different for the τ measures. However, τ_{ind} appears to increase faster with P than τ_{joint} does. Therefore the gain in the predictive accuracy for ranking problems is larger for the individuals with fewer preference information pieces ($v_{003} = 0.064$ for $\delta\tau$). This means that the individuals with poorer preference information benefit more from accounting for the structural patterns in the global shape of value functions. The value function for an individual with fewer pairwise comparisons is subjected to a smaller number of constraints and entails more degrees of freedom. Thus, incorporating information on the structural patterns yields a greater improvement in the predictive accuracy of these constructed value function. For choice problems, the differences in predicted values for the random parameter related to the groups distinguished by different amount of preference information are negligible between the independent and joint preference estimation methods. The latter is consistent with our observation from the graphical analysis. In fact, by allowing a variation in the intercepts by groups with different amount of preference information, the BIC measure for $\delta\alpha$ did not decrease (it increased from -119005 to -118999). Finally, after controlling the sources of variation by individuals and by groups with different numbers of pairwise comparisons, it can be seen that η has a significantly positive impact on the predictive validity measures, demonstrating a logarithmic pattern.

Overall, the results for both ranking and choice problems provide no evidence for supporting the hypothesis that the decision makers who are under-represented in terms of the input preference information to the joint preference estimation method are affected negatively in terms of the predictive validity of their constructed value functions.

Technical details: The simulations, sampling and preference estimations were implemented in MATLAB $R2017a$ package, and computations were performed on OS X v10.9.4 Mac PC with 2,93 GHz Intel core i7 processor and 16GB memory. The total execution time was 59.5, 4, and 42 hours for the Simulation Analysis, Study 1, and Study 2, respectively.

6 Conclusions

Results from empirical research and experimental analysis in the literature attributes global shape of a value function to the higher-order factors underlying the decision problem, such as organizational behavior (Pennings and Smidts, 2003), decision environment (Pennings and Garcia, 2009), and contextual factors (Hershey et al., 1982). In this paper, we introduced an analytical framework for a joint estimation of preferences of a group of decision makers by accounting for structural patterns in global shape of value functions.

The proposed framework, based on convex optimization, aims at constructing value functions for a group of decision makers, without making any assumption on the shape of value function or uniformity of general shapes across the decision makers, by capturing the structural patterns that regulate a value function's general shape, and by incorporating them in the preference estimation process in a unified framework. The method compares pairwise the regions of a performance scale in terms of slope deviations of value functions from the group's mean. Since the comparisons are made both across all pairs of regions in search of the systematic interdependencies, and for all individual regions in search of uniformity in sensitivities, the framework captures general shapes of value functions. Moreover, regularities in the local shape, i.e. curvature, is captured through comparing the neighbor regions of value functions.

Let us summarize the main findings discussed in the paper:

- We found that – when dealing with ranking problems – accounting for the structural patterns in the global shape of value function always improves the predictive validity, and the improvement is independent from the individual characteristics of decision makers. For choice problems, however, even if the improvement in the predictive accuracy at the aggregate level is positive, the impact demonstrates large variation across the individuals. Analysis of the preferences from real decision makers with unbalanced preference information revealed that for 49% of individuals such an impact was increasing with η, for 26% it followed an inverted U shape, and for the rest it was decreasing.

- We confirmed that the improvement in the predictive validity for ranking problems was particularly larger when smaller amount of preference information was available (confirmed by both simulation analysis and Study 1), or for decision makers with relatively smaller number of pairwise comparisons compared to the population (confirmed by Study 2). This is intuitive, because in an extreme case where excessively large amount of preference information is supplied by the decision makers, no further information – including that on structural patterns – is needed to improve the predictive accuracy, and an independent estimation method for construction of value functions should work well.

- We found that for choice problems, improvement in the predictive accuracy by accounting for structural patterns is independent from the amount of preference information, and it is larger when greater number of decision makers are involved.

These findings are confirmed by an extensive simulation analysis, as well as analysis of preferences of real decision makers with different levels of available preference information and various positions with respect to their contribution to an overall supplied preference information.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46(11):1497–1512.

Abdellaoui, M., Bleichrodt, H., and Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, 53(10):1659–1674.

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using eigen and s4. *R package version*, 1(7):1–23.

Bleichrodt, H. and Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science*, 46(11):1485–1496.

Ciomek, K., Kadziński, M., and Tervonen, T. (2017). Heuristics for prioritizing pair-wise elicitation questions with additive multi-attribute value models. *Omega*, 71:27 – 45.

Fishburn, P. C. and Kochenberger, G. A. (1979). Two-piece von neumann-morgenstern utility functions. *Decision Sciences*, 10(4):503–518.

Friedman, M. and Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, 56(4):279–304.

Ghaderi, M. (2017). *Preference Disaggregation: Towards an Integrated Framework*. PhD thesis, ESADE Business School.

Ghaderi, M., Ruiz, F., and Agell, N. (2017). A linear programming approach for learning non-monotonic additive value functions in multiple criteria decision aiding. *European Journal of Operational Research*, 259(3):1073–1084.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38(1):129–166.

Greco, S., Mousseau, V., and Słowiński, R. (2008). Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191(2):416–436.

Green, P. E. and Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 5(2):103–123.

Green, P. E. and Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *The Journal of Marketing*, pages 3–19.

Hershey, J. C., Kunreuther, H. C., and Schoemaker, P. J. (1982). Sources of bias in assessment procedures for utility functions. *Management Science*, 28(8):936–954.

Hox, J. J., Moerbeek, M., and Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Kadziński, M., Ghaderi, M., Wasikowski, J., and Agell, N. (2017). Expressiveness and robustness measures for the evaluation of an additive value function in multiple criteria preference disaggregation methods: an experimental analysis. *Computers & Operations Research*, 87:146–164.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Keeney, R. L. (1996). *Value-focused thinking*. Harvard University Press.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

Kilka, M. and Weber, M. (2001). What determines the shape of the probability weighting function under uncertainty? *Management Science*, 47(12):1712–1726.

Luke, D. A. (2004). *Multilevel modeling*, volume 143. Sage.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Pennings, J. M. and Garcia, P. (2009). The informational content of the shape of utility functions: financial strategic behavior. *Managerial and Decision Economics*, 30(2):83–90.

Pennings, J. M. and Smidts, A. (2000). Assessing the construct validity of risk attitude. *Management Science*, 46(10):1337–1348.

Pennings, J. M. and Smidts, A. (2003). The shape of utility functions and organizational behavior. *Management Science*, 49(9):1251–1263.

Sándor, Z. and Wedel, M. (2005). Heterogeneous conjoint choice designs. *Journal of Marketing Research*, 42(2):210–218.

Schunk, D. and Betsch, C. (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology*, 27(3):386–401.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Snijders, T. A. (2005). Power and sample size in multilevel modeling. *Encyclopedia of Statistics in Behavioral Science*, 3:1570–1573.

Trepel, C., Fox, C. R., and Poldrack, R. A. (2005). Prospect theory on the brain? toward a cognitive neuroscience of decision under risk. *Cognitive Brain Research*, 23(1):34–50.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Tversky, A. and Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica: Journal of the Econometric Society*, pages 1255–1280.

Appendices

A Graphical representation of the experimental results for Study 1

Figures 21 and 22 represent how predictive accuracy measures in Study 1 vary with η across individual decision makers.

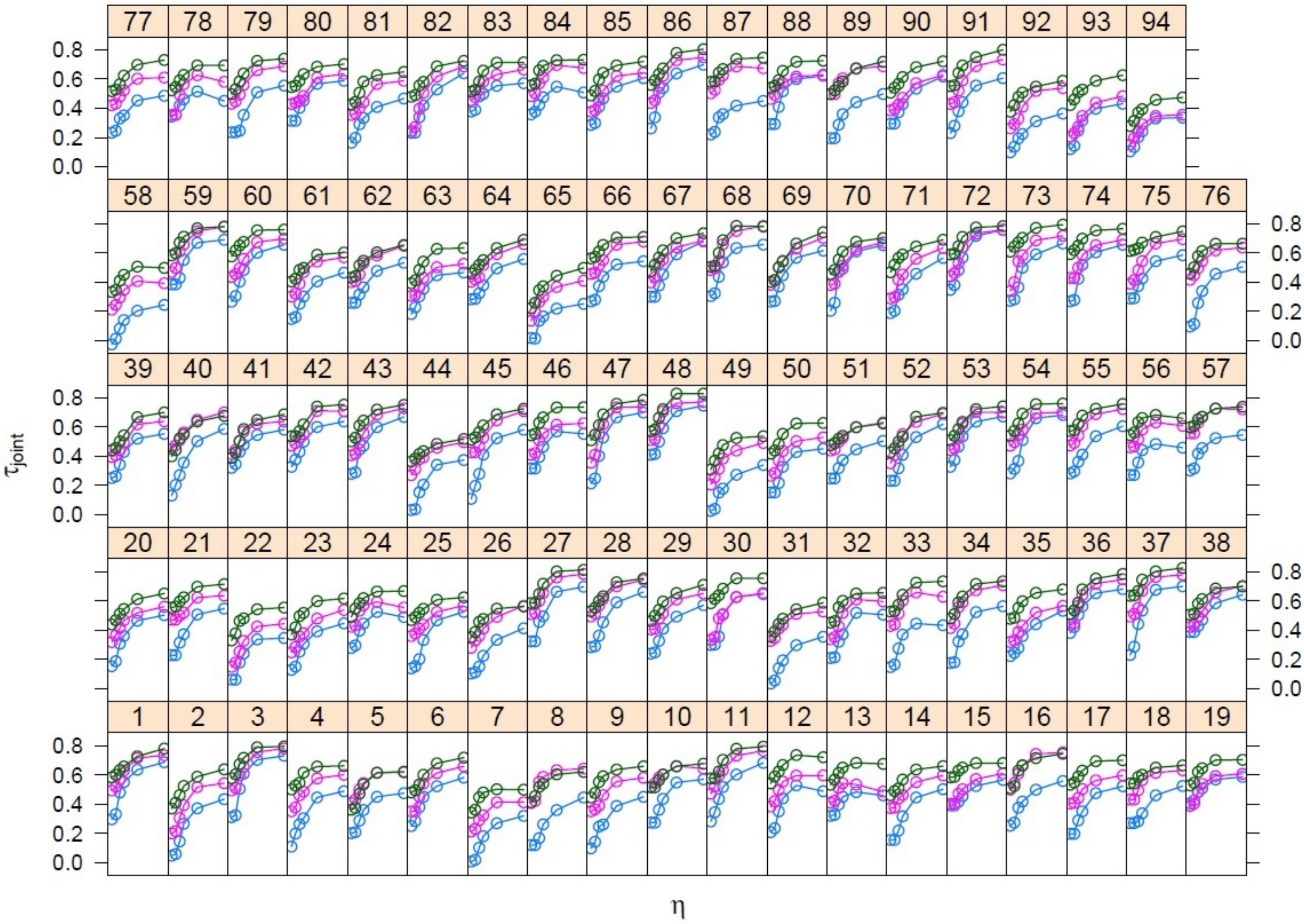

Figure 21: Relation between τ_{joint} and η for the $R = 94$ decision makers for different levels of P (blue – 3; purple – 5; green – 7) in Study 1.

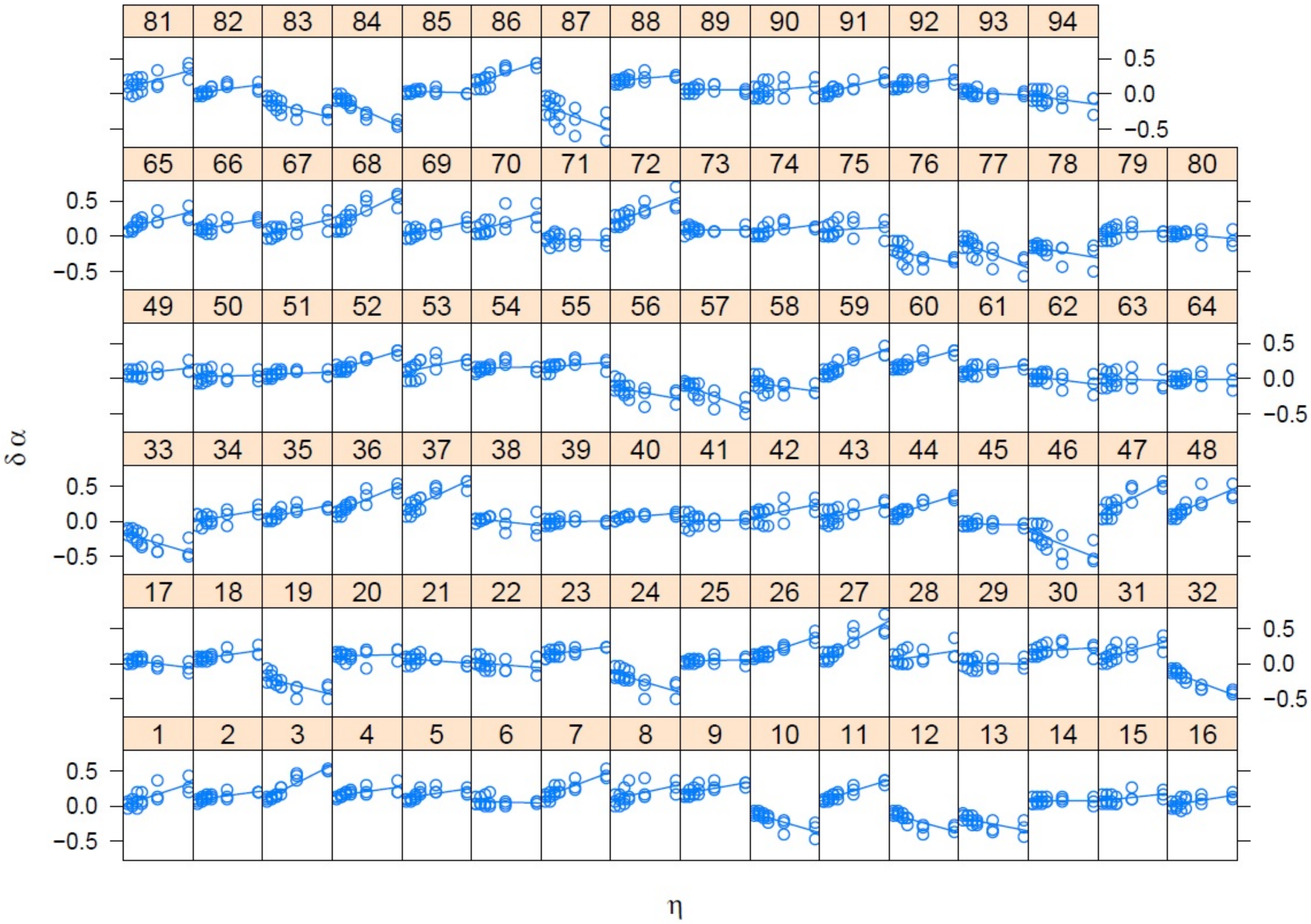

Figure 22: Relation between $\delta\alpha$ and η for the $R = 94$ decision makers in Study 1.

B Graphical representation of the experimental results for Study 2

Figures 23 to 28 represent variations in predictive accuracy measures in Study 2 across the individual decision makers; Figures 29, 30, and 31 represent such variations in the expressiveness measures; Figure 32 represents variations in the predictive accuracy and expressiveness measures by the groups defined in Study 2 according to the relative preference information; and Figures 33 to 36 represent how predictive accuracy measures in Study 2 vary with η across individual decision makers.

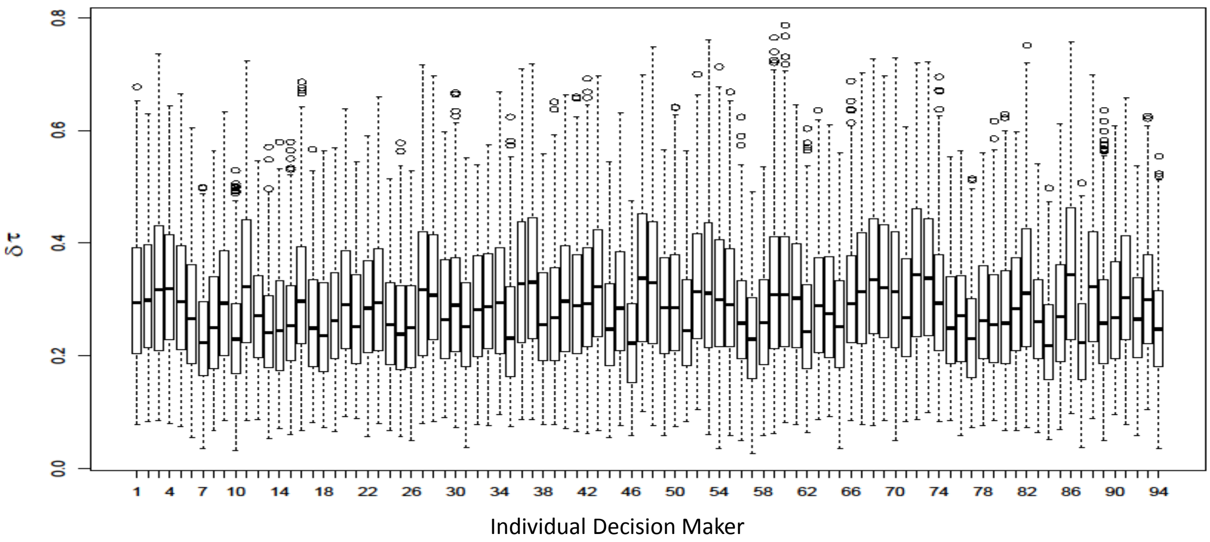

Figure 23: Variation of $\delta\tau$ by individuals in Study 2.

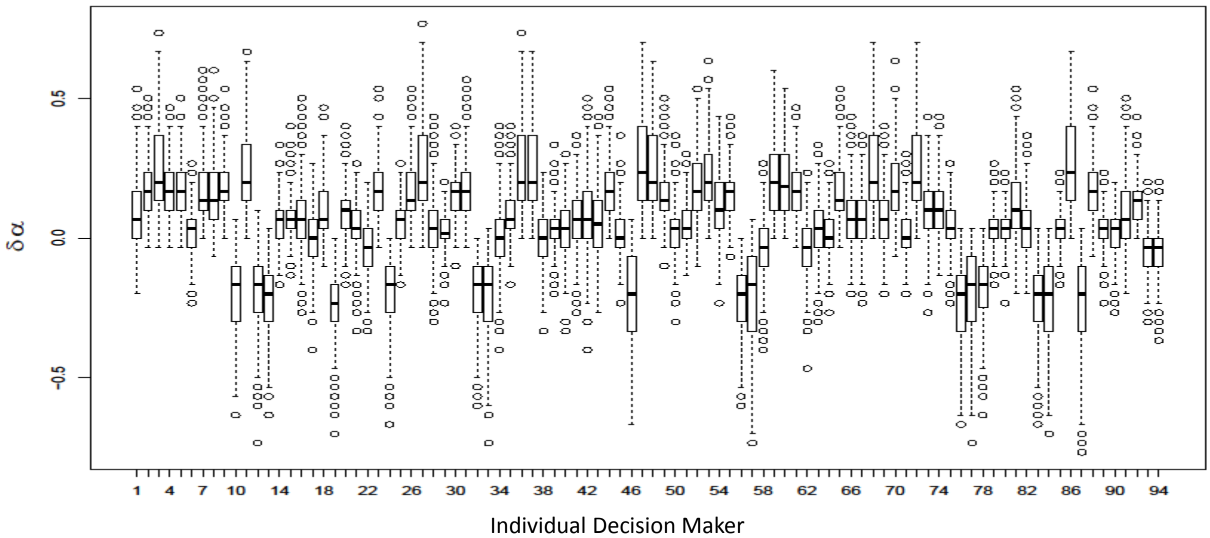

Figure 24: Variation of $\delta\alpha$ by individuals in Study 2.

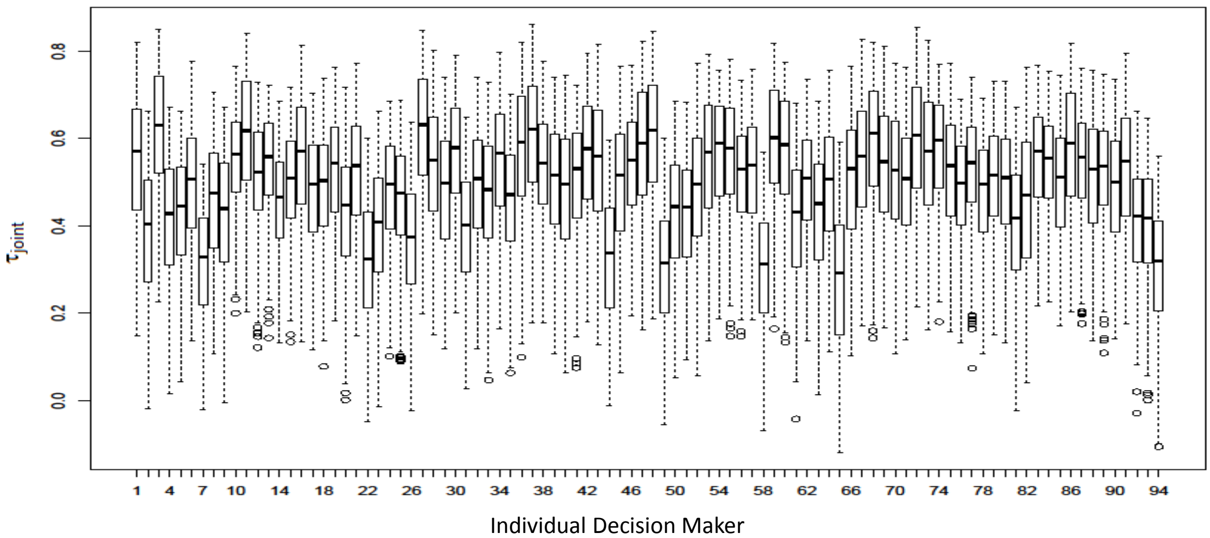

Figure 25: Variation of τ_{joint} by individuals in Study 2.

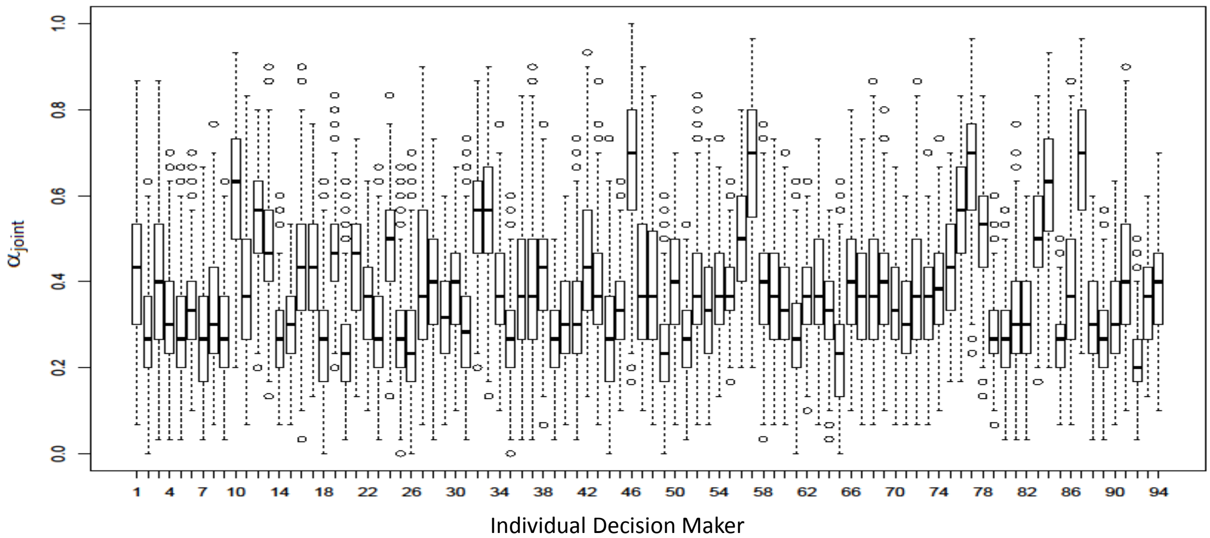

Figure 26: Variation of α_{joint} by individuals in Study 2.

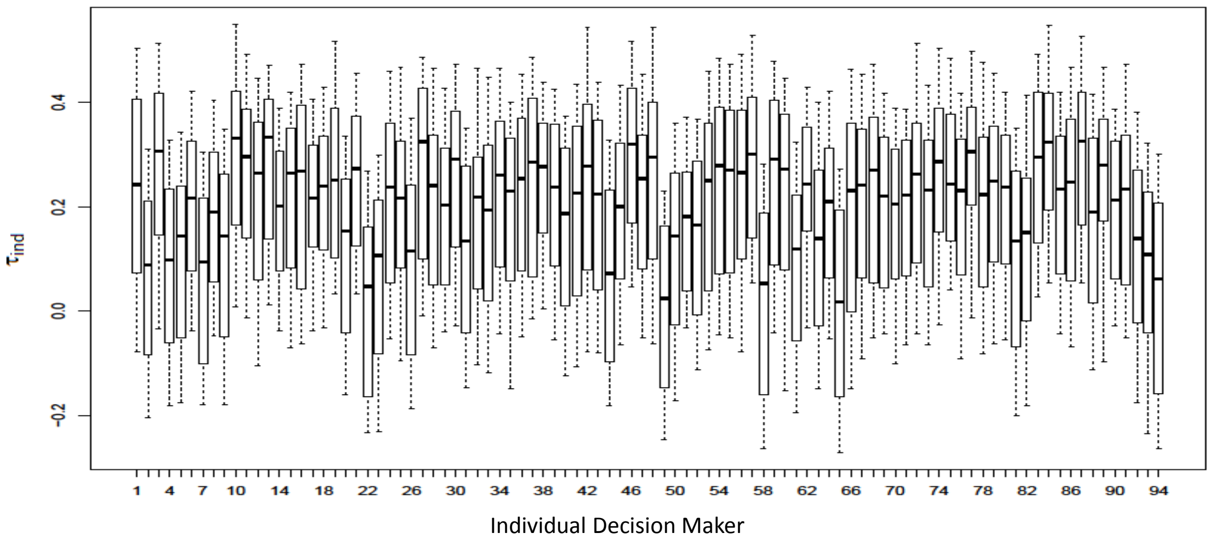

Figure 27: Variation of τ_{ind} by individuals in Study 2.

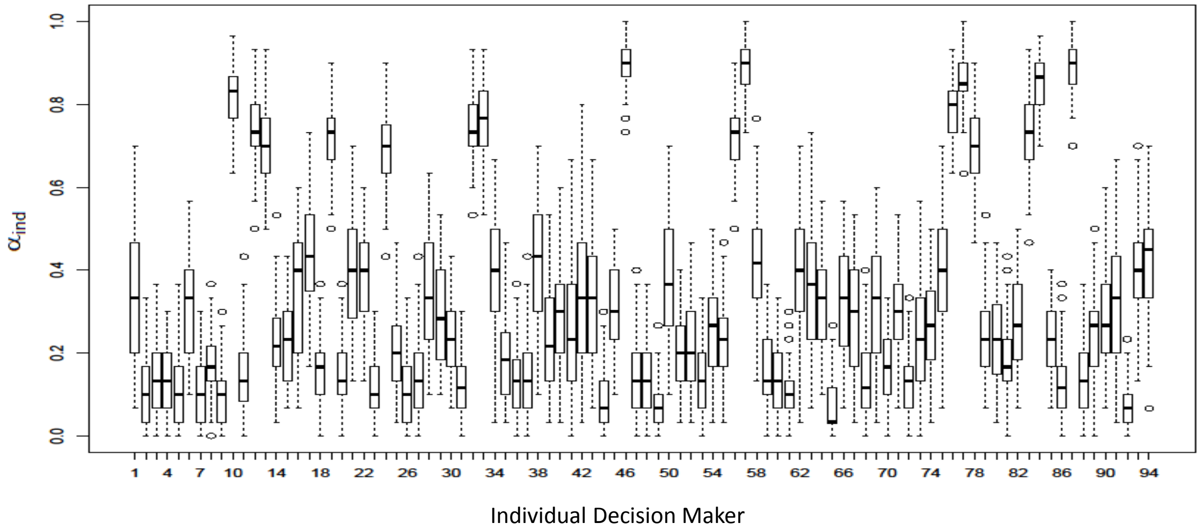

Figure 28: Variation of α_{ind} by individuals in Study 2.

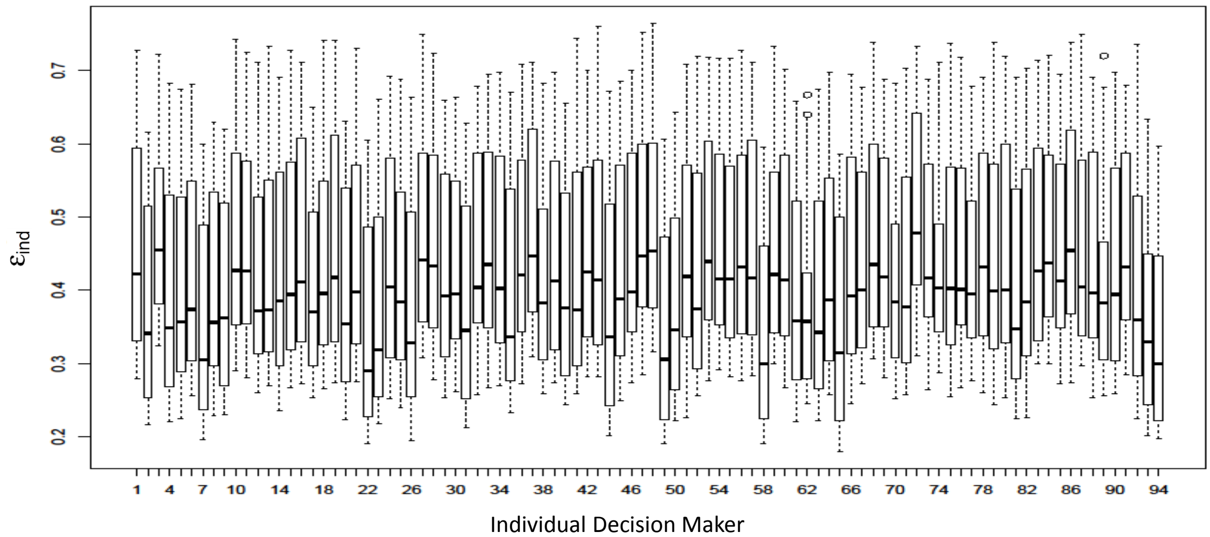

Figure 29: Variation of ε_{ind} by individuals in Study 2.

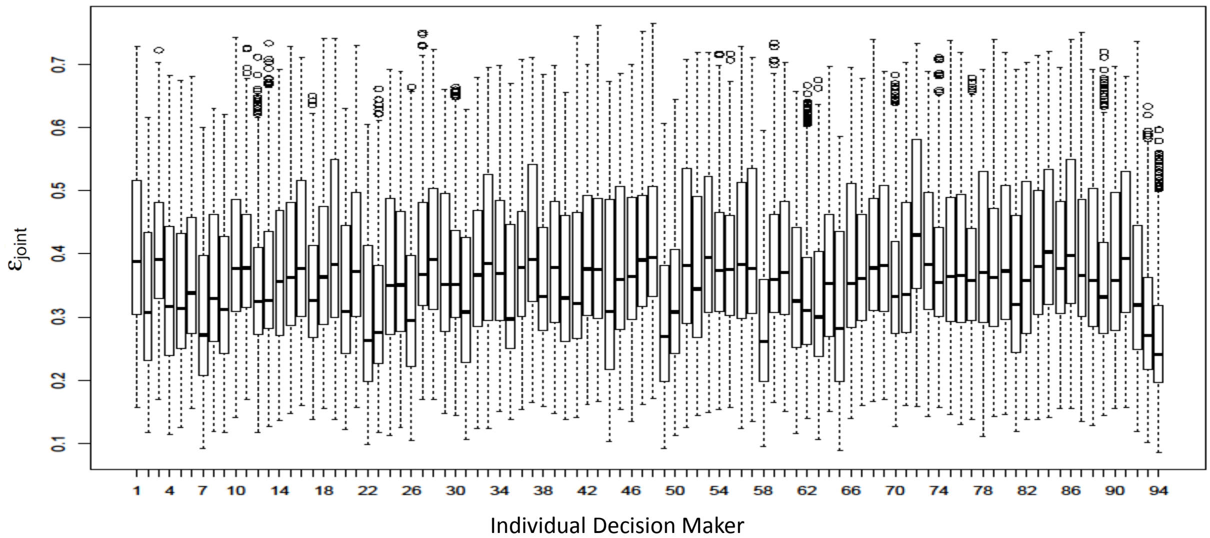

Figure 30: Variation of ε_{joint} by individuals in Study 2.

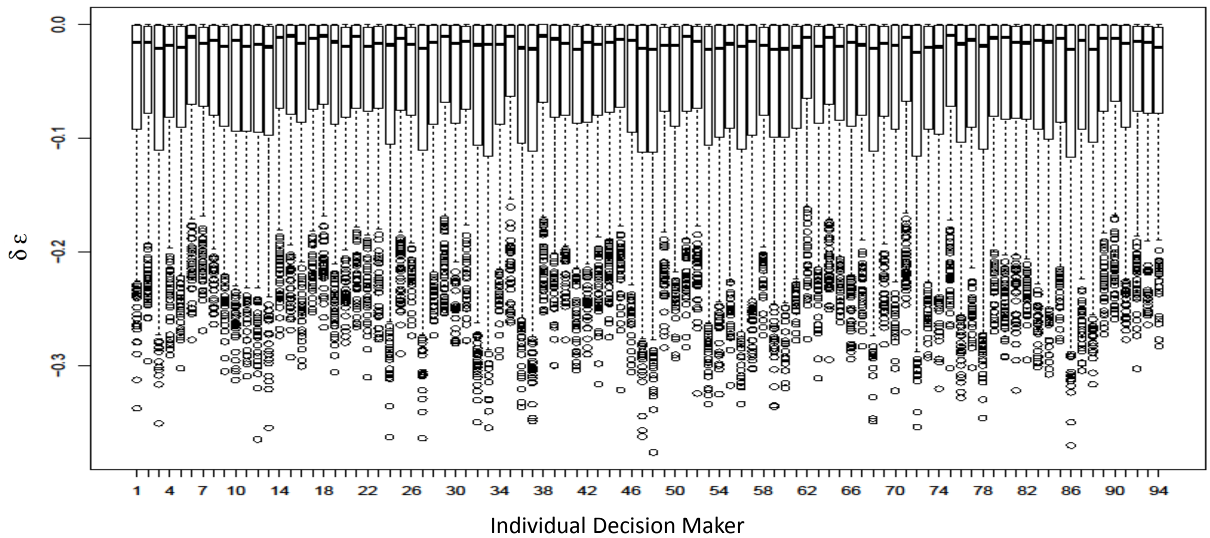

Figure 31: Variation of $\delta\varepsilon$ by individuals in Study 2.

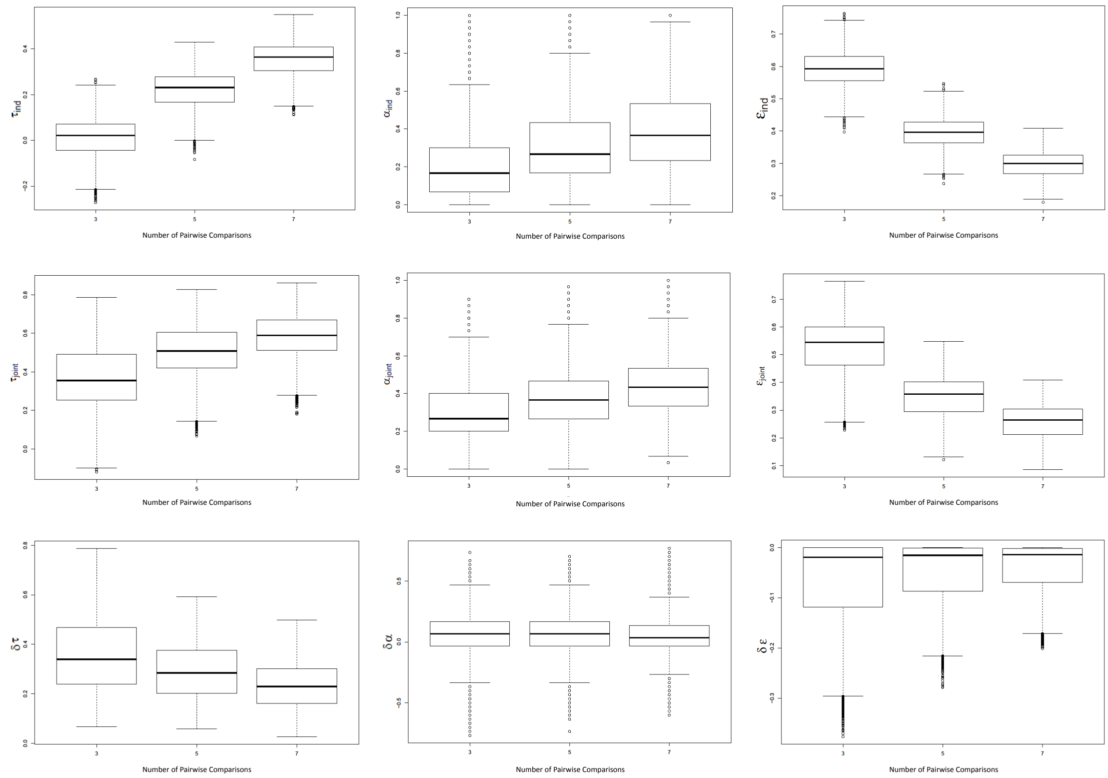

Figure 32: Variation of τ (left), α (middle) and ε (right) measures in Study 2 obtained by employing the independent estimation method (top), joint estimation method (middle), and variation in differences of such measures obtained by the two methods (bottom) by levels of P (3, 5, 7, from the left to the right side in each box, respectively).

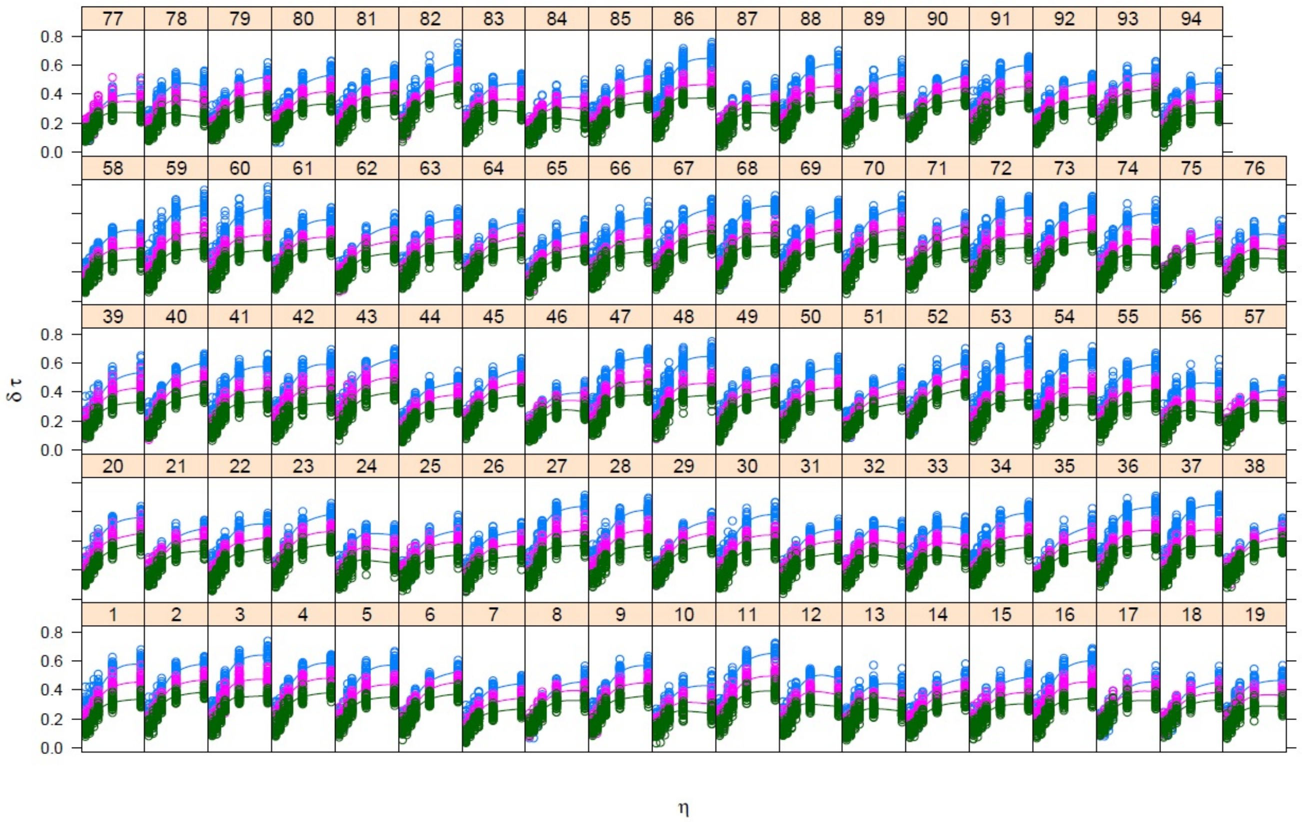

Figure 33: Relation between $\delta\tau$ and η for the $R = 94$ decision makers for different levels of P (blue – 3; purple – 5; green – 7) in Study 2.

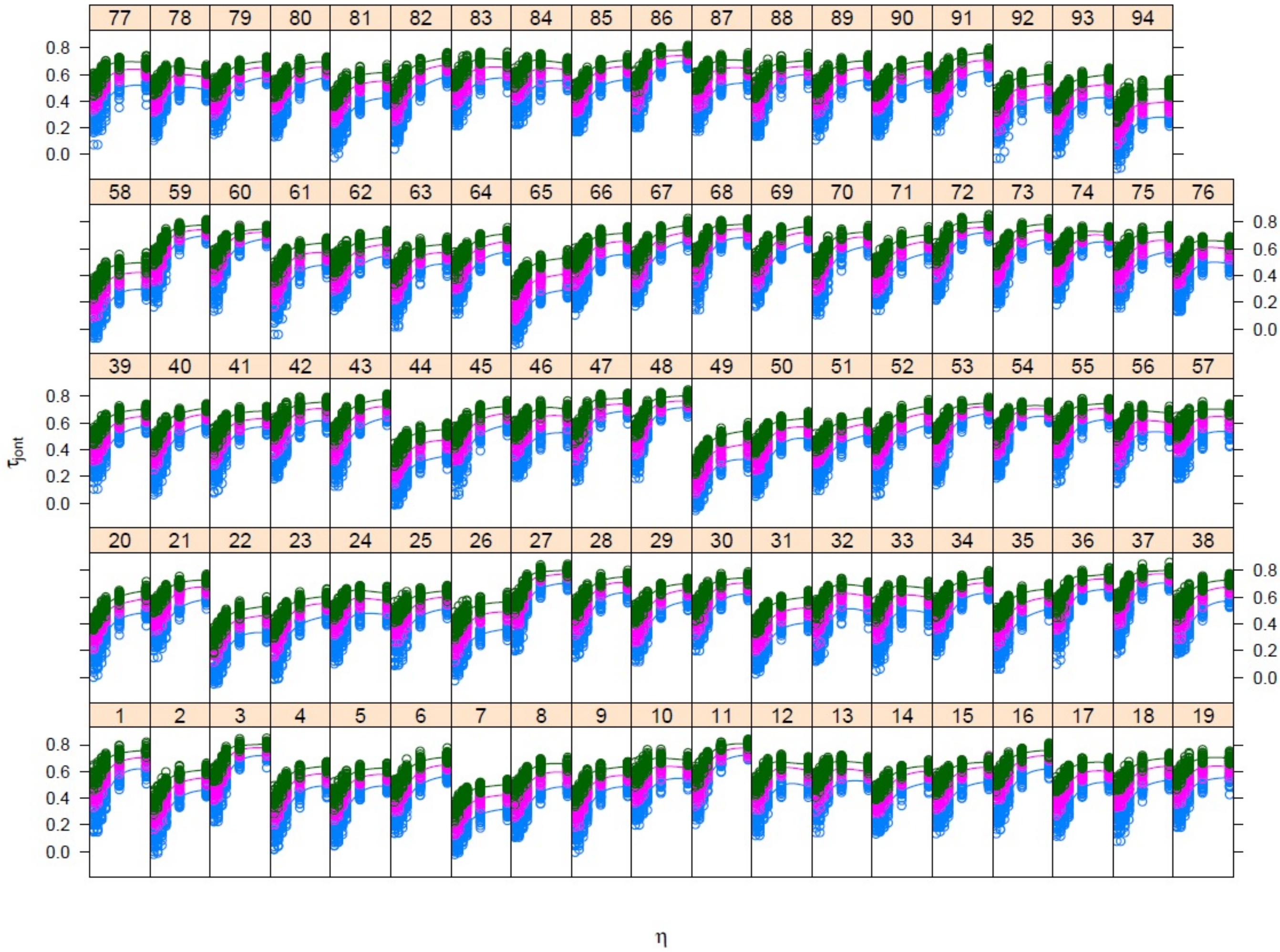

Figure 34: Relation between τ_{joint} and η for the $R = 94$ decision makers for different levels of P (blue – 3; purple – 5; green – 7) in Study 2.

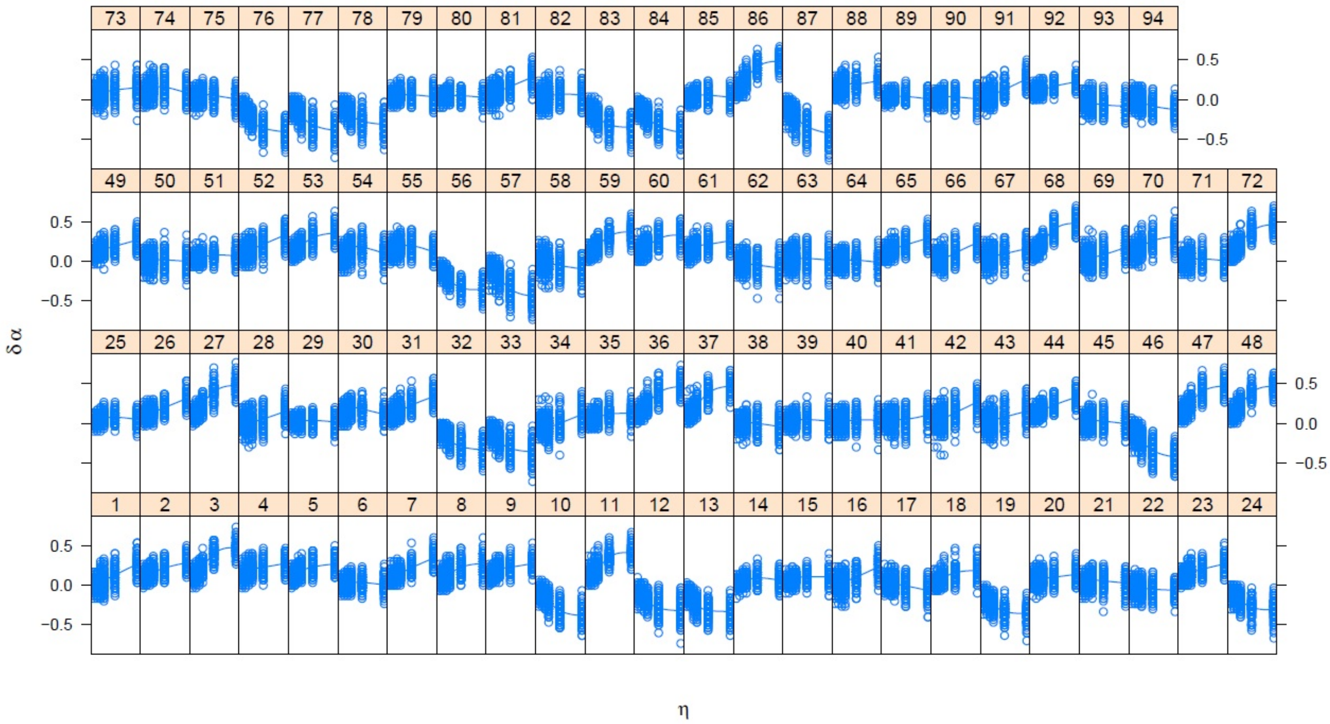

Figure 35: Relation between $\delta\alpha$ and η for the $R = 94$ decision makers in Study 2.

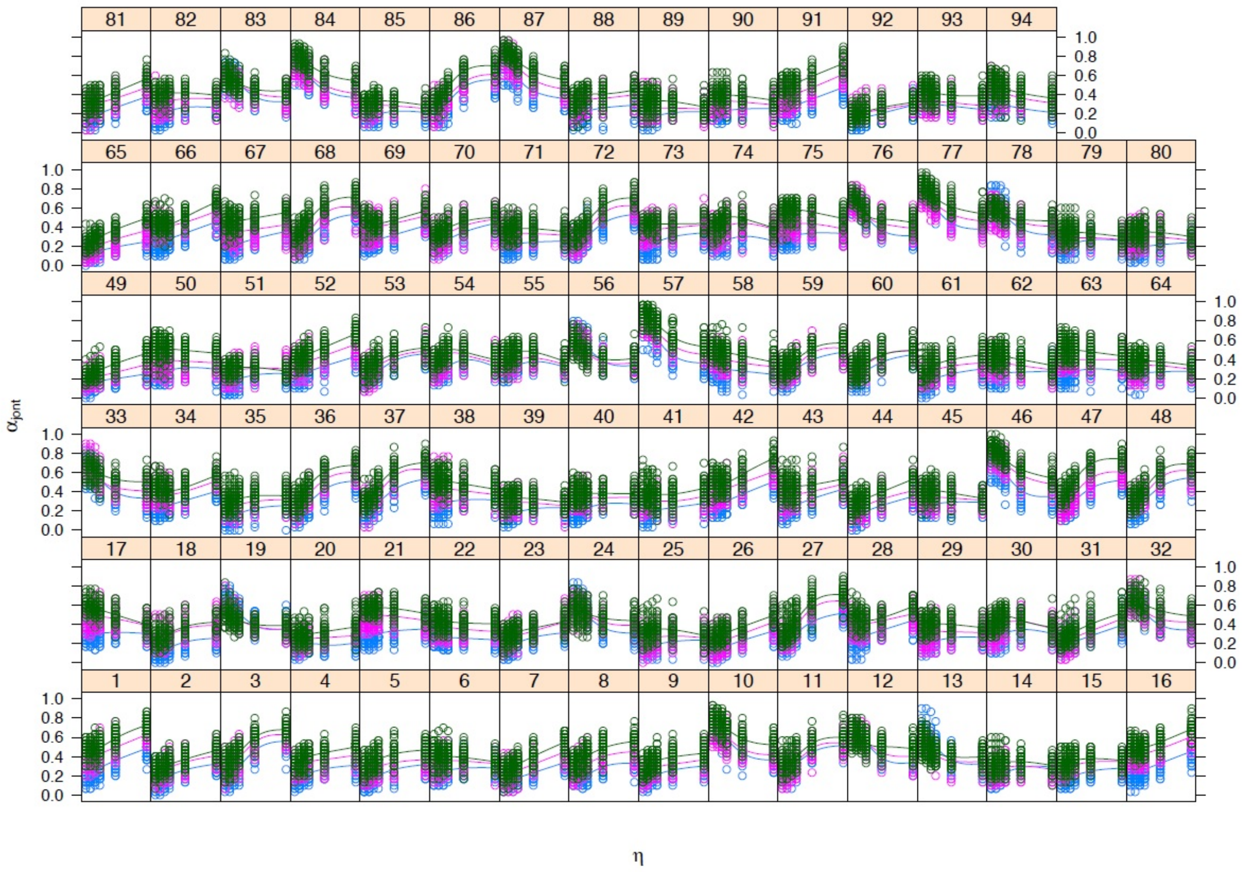

Figure 36: Relation between α_{joint} and η for the $R = 94$ decision makers for different levels of P (blue – 3; purple – 5; green – 7) in Study 2.