

## TOWARDS A FRAMEWORK TO COMBINE MULTIOBJECTIVE OPTIMIZATION AND ECONOMETRICS AND AN APPLICATION IN ECONOMICS OF EDUCATION

MARIANO LUQUE<sup>1</sup>, OSCAR D. MARCENARO-GUTIERREZ<sup>2</sup>,  
SANDRA GONZÁLEZ-GALLARDO<sup>1,3,\*</sup>  AND ANA B. RUIZ<sup>1</sup>

**Abstract.** In this paper, we propose a theoretical framework that combines econometric and multi-objective programming methodologies to help researchers to identify and achieve optimal solutions to socio-economic and management problems. Sometimes, it is important to analyse which combination of values of the explanatory variables -in an econometric model- would imply the simultaneous achievement of the best values of the response variables. In such situations, if certain degree of conflict is observed among the response variables, we propose to formulate a multiobjective optimization problem based on the conclusions obtained from a regression analysis. Subsequently, the application of multiobjective optimization techniques allows gaining a better insight about the conflicting relation between the response variables, and how a balanced “optimal” situation among them could be achieved. This piece of information can be hardly extracted just by econometric techniques. An application in the field of economics of education, related to the analysis of the students’ well-being as a way to improve their academic performance, demonstrates the potential of our proposal.

**Mathematics Subject Classification.** 90C29, 62P20, 62P25.

Received October 14, 2021. Accepted May 24, 2022.

### 1. INTRODUCTION

Broadly speaking, *econometrics* aims at giving empirical quantitative content to economic models, in order to forecast how some variables would be affected depending on the changes in some others. In other words, we estimate econometric models to quantify and verify predictions from economic theories -and other related fields- by means of regression analysis, which studies the relationship existing between the (endogenous) response variable/s and the (exogenous) explanatory variable/s. When the phenomenon studied is analysed through several response variables, the econometric analysis may reveal the existence of certain degree of conflict among them, in the sense that the most desired values observed in the data for some of the response variables may be associated only with a worsening of others, and *vice versa*. This can be observed, *e.g.* by means of the

---

*Keywords.* Multiple objective decision-making, econometrics, multiobjective optimization, economics of education.

<sup>1</sup> Department of Applied Economics (Mathematics), Universidad de Málaga, C/ Ejido S/N, Málaga 29071, Spain.

<sup>2</sup> Department of Applied Economics (Statistics and Econometrics), Universidad de Málaga, C/ Ejido S/N, Málaga 29071, Spain.

<sup>3</sup> Department of Quantitative Methods for Business and Economics, Universidad de Granada, P.º de Cartuja, 7, Granada 18011, Spain.

\*Corresponding author: [sandragg@uma.es](mailto:sandragg@uma.es)

descriptive statistics of the response variables. However, this type of analysis cannot give much more insights than the conflicting nature of the variables, to the extent that it cannot explain how to attain “optimal” values simultaneously for all the response variables of the phenomenon under research. In such a situation, the application of multiobjective optimization techniques in conjunction with econometric techniques can enable us to obtain very valuable information.

*Multiobjective optimization* is intended at solving problems arising in fields such as *e.g.* economics, management, industry or engineering, in which several conflicting criteria (modelled through objective functions) must be optimized (maximized or minimized). The feasible set of alternatives (solutions) is determined by the set of constraints modelling the situation studied taking into account the observed data. Because of the conflicting nature of the criteria, it is not possible “in most of the cases” to find a single solution at which all the objectives are optimized simultaneously and, thus, the so-called Pareto optimal or efficient solutions are identified. At these solutions, no objective function can be improved without worsening, at least, one of the others.

Under this scope, in this paper, we propose a theoretical framework that combines econometric and multiobjective optimization techniques to help researchers and/or decision makers to have a more precise view and a better knowledge of the phenomenon under scrutiny. The framework suggested involves two stages. Firstly, an econometric analysis has to be conducted in order to find and estimate relationships between the explanatory and the response variables considered. These relations can be obtained by means of different models, such as linear, polynomial, exponential, etc. It is important to detect if a conflict exists among the response variables in this econometric study, since this will support the analysis of the problem from a multiobjective optimization point of view. Secondly, we suggest building a multiobjective optimization problem from the econometric analysis previously done, making use of the correlations found among the variables. To be more precise, the regressions and correlations are used to formulate the objective functions and to delimit the set of realistic values for the explanatory variables through constraints. The idea is to gain some knowledge about the conflicting relation existing among the response variables. By applying different multiobjective optimization techniques, we can reach some findings about the type of solutions that are feasible, about the sacrifices among the objectives (trade-offs) needed to attain a desirable “optimal” solution, and so on.

The main purpose of the framework proposed is to allow practitioners and decision makers<sup>1</sup> to obtain further information about the values of the explanatory variables that would permit to achieve compromise optimal values of the response variables simultaneously in the future. These optimal values will be obtained based on the expectations of the decision makers about a hypothetical desirable situation regarding the phenomenon under scrutiny. In particular, the conclusions extracted can be used to disentangle to which extent we can determine strategies that make the explanatory variables reach certain values improving the response variables as much as possible. This can be of great interest in several situations. For example, when researchers desire to know if there exists a combination of values for the explanatory variables that would make all the response variables to achieve their “best possible” values at the same time. Besides, if they want to investigate how to influence the explanatory variables to make the response ones to attain certain specific values that are desirable and relevant according to the data observed. In addition, this framework is useful to investigate how sensitive these “optimal” response variable values may be to small variations in the explanatory variable values. Furthermore, we can foresee the impact of giving more importance to the improvement of some of the response variables, having information about the sacrifice required in the values achieved by the others. Note that these findings can be hardly obtained just by econometric techniques.

To the best of our knowledge, few previous works have considered a similar mixed econometric-multiobjective optimization procedure. Spivey and Tamura [27] combined econometrics and the multiobjective optimization technique known as goal programming for dealing with problems of policy formulation and decision making, but just a preference function was optimized subject to a set of econometric constraints. In [30], a multiobjective optimization method to study an econometric model of macroeconomics policy in Finland was applied. In [20], a linear mixed integer multiobjective optimization problem was built from an econometric model to study the

---

<sup>1</sup>In the context of this paper, we refer by *decision maker* to the analyst, researcher, econometrician or policy maker that is interested in analyzing a socio-economic problem from both the econometric and the multiobjective optimization points of view.

satisfaction level of Spanish workers and they identified the profile of an “ideal” (optimal) satisfied Spanish worker. In [18], a multiobjective optimization approach was obtained from the econometric analysis of the education outcomes of students, with the aim of researching the educational inputs that may be affected to achieve an optimum balanced performance of students in reading, mathematics and English courses. Also, [21] investigated the potential balance between some teacher characteristics, particularly teachers’ satisfaction and different measures of their pupils’ performance, in order to optimize some Spanish educational system outputs. More recently, in [11], a multiobjective interval model was proposed, based on the results of an econometric estimation, to explore the trade-offs among different dimensions of job satisfaction. González-Fernández *et al.* [10] used a similar procedure, involving an econometric analysis and an application of goal programming, to study the profile of the most profitable insurers. Later, [12] applied interval multiobjective optimization techniques jointly to regression analysis to shed some light about the compromises of specific aspects of workers’ personal and working conditions in different scenarios.

As said, there are several works available in the literature that apply a mixed econometric and multiobjective optimization methodology. However, the aforementioned articles were focused on specific socio-economic problems and, therefore, applied this methodology ad-hoc for each specific problem. Instead, in this paper, we generalize the combination of both techniques to a framework for a generic socio-economic problem and we identify general features and steps to be followed to combine both techniques. Indeed, our theoretical framework enables a more flexible and general formulation than the models considered in these previous works. In addition to formulating the linear case theoretically, the non-linear case is also considered in our proposal, allowing the use of non-linear regressions if needed. We also go a step further in the linear case and we formulate more realistic models that consider confidence intervals for the coefficients of the objective functions.

In order to show the potential of our framework, we describe an illustrative example related to economics of education. Specifically, we apply our proposal to a problem aimed at analysing the levels of well-being of students in secondary school by means of a set of indicators. The purpose is to identify the factors of the educational context that would enable to achieve the best possible levels of these indicators in the future, in order to formulate educational policies that promote an increase of the students’ well-being, as a way to enhance their academic performance. In this example, a reference point-based methodology is applied to solve the multiobjective optimization problem built from the econometric analysis carried out, and interesting findings are concluded.

The rest of this paper is organised as follows. Section 2 provides the basic concepts and notations of econometrics and multiobjective optimization. The theoretical framework proposed is described in Section 3. Next, Section 4 discusses some multiobjective optimization techniques that can be used for solving the resulting problem. We demonstrate the potential of our proposal using the illustrative example about the students’ well-being in Section 5. Finally, the main contributions of this paper are summarized in Section 6.

## 2. BACKGROUND

### 2.1. Econometrics

Let us assume that we are studying a phenomenon and we want to find dependence relations of a set of response variables, denoted by  $y_j$  for  $j = 1, \dots, m$ , with respect to a set of explanatory variables, referred to as  $x_i$  for  $i = 1, \dots, k$ . For a sample of  $n$  observational data, each vector of explanatory variables is associated with a vector of response variables as follows:

$$(x_1(r), \dots, x_k(r)) \rightarrow (y_1(r), \dots, y_m(r)), \quad \text{for } r = 1, \dots, n.$$

In order to study the dependency among them, an econometric analysis can be carried out by estimating a regression model in which the response variables are regressed on the explanatory variables. Let us consider a general regression model formulated as follows:

$$y_j = \hat{y}_j(x_1, \dots, x_k) + \epsilon_j, \quad \text{for } j = 1, \dots, m, \quad (2.1)$$

where  $\hat{y}_j$  is the predicted value of the response variable  $y_j$ , which is a function of the explanatory variables  $x_1, \dots, x_k$ , and  $\epsilon_j$  is a random disturbance (inherently unobservable and generally assumed to be normally distributed). Observe that the parameter  $r$ , which denotes the observation number, has been deleted in the variables  $x_1, \dots, x_k$  and  $y_1, \dots, y_m$  to simplify the notation. For  $j = 1, \dots, m$ , the functions  $\hat{y}_j$ , referred to as predictors of the response variables, can be *e.g.* linear, polynomial, exponential, etc., depending on the nature of the data and the model considered. For example, in the linear case, they are given by the following formulation:

$$\hat{y}_j = \hat{\alpha}^j + \hat{\beta}_1^j \cdot x_1 + \dots + \hat{\beta}_k^j \cdot x_k, \quad \text{for } j = 1, \dots, m,$$

where  $\hat{\beta}^j = (\hat{\beta}_1^j, \dots, \hat{\beta}_k^j)^T$  is the vector of regression coefficients (slopes) and  $\hat{\alpha}^j$  is an estimated population intercept. The regression coefficients can be estimated by means of any suitable methodology, such as *e.g.* ordinary least squares regression.

## 2.2. Multiobjective optimization

In general, a multiobjective optimization problem [9, 23] can be formulated as:

$$\begin{aligned} \max \quad & \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\} \\ \text{subject to } & \mathbf{x} \in X, \end{aligned} \tag{2.2}$$

where  $f_j : \mathbb{R}^k \rightarrow \mathbb{R}$  con  $j = 1, \dots, m$  are the  $m$  (with  $m \geq 2$ ) conflicting objective functions to be optimized simultaneously over the feasible set  $X \in \mathbb{R}^k$ , constituted by the feasible decision vectors  $\mathbf{x} = (x_1, \dots, x_k)^T$ . Their images in the objective space, given by  $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^T$  for any  $\mathbf{x} \in X$ , are referred to as objective vectors and form the so-called feasible objective region  $Z = \mathbf{f}(X) \in \mathbb{R}^m$ .

Finding a single solution optimizing all the criteria at the same time is usually impossible in most of the cases because of the conflict degree among the objectives, but the so-called efficient or Pareto optimal solutions do exist instead. A decision vector  $\mathbf{x}^0 \in \mathbb{R}$  is said to be efficient or Pareto optimal for problem (2.2) if there does not exist any other vector  $\mathbf{x} \in X$  such that  $f_j(\mathbf{x}^0) \leq f_j(\mathbf{x})$  for every  $j = 1, \dots, m$ , and  $f_i(\mathbf{x}^0) < f_i(\mathbf{x})$  for at least one index  $i$ . The corresponding objective vector  $\mathbf{f}(\mathbf{x}^0)$  is called a Pareto optimal objective vector. All efficient solutions form the Pareto optimal set in the decision space ( $E$ ) and the Pareto optimal front in the objective space ( $\mathbf{f}(E)$ ).

Usually, there exist more than one Pareto optimal solution, so it is useful to know the ranges of the objective functions in the Pareto optimal front. On the one hand, the lower bounds are set by the nadir vector  $\mathbf{z}^{\text{nadir}} = (z_1^{\text{nadir}}, \dots, z_m^{\text{nadir}})^T$ , where  $z_j^{\text{nadir}} = \min_{\mathbf{x} \in E} f_j(\mathbf{x})$  for all  $j = 1, \dots, m$ , while the upper bounds are given by the ideal vector  $\mathbf{z}^* = (z_1^*, \dots, z_m^*)^T$ , where  $z_j^* = \max_{\mathbf{x} \in S} f_j(\mathbf{x}) = \max_{\mathbf{x} \in E} f_j(\mathbf{x})$  for all  $j = 1, \dots, m$ . In practice, the nadir point is usually approximated since its computation is difficult as the set  $E$  is unknown [14]. Both the ideal and nadir vectors are frequently used to normalize the objective functions (*e.g.* by dividing each objective by the difference of its corresponding ideal and nadir values).

## 3. A THEORETICAL ECONOMETRIC-MULTIOBJECTIVE OPTIMIZATION FRAMEWORK

By studying the regression model, noteworthy relations and dependencies between the different explanatory and response variables can be identified, which in turn may provide some interesting conclusions about the phenomenon under research. But, in econometrics, sometimes it is valuable to discuss and analyse how “the best values” for the response variables could be achieved simultaneously, if possible, considering all the feasible values that the explanatory variables can take (not only the values observed in the sample). This is the main purpose of the combined methodology proposed here. The information obtained in this way may enable supporting bad or good decisions taken regarding the phenomenon under study, by the identification of key factors concerning the explanatory variable values that would allow the simultaneous achievement of such desired “optimal” response variable values.

To shed some light in this regard, if the data indicate the existence of certain degree of conflict among the response variables, we can formulate a multiobjective optimization problem using the econometric estimations. The conflicting nature of the response variables can be revealed by an analysis of their descriptive statistics, which may evidence the achievement of potential good values for some variables at the expense of a sacrifice in the others. Once the multiobjective optimization problem is built, interesting findings can be extracted depending on the purpose of the study. Actually, decision makers -researchers or politicians involved in the study- can indicate their preferences about certain features of the phenomena studied, which can be used to determine Pareto optimal solutions fitting “as much as possible” the expectations expressed through these preferences. Furthermore, information about which policies should be followed in order to achieve certain goals -in the future- for the response variables can be provided to the decision makers, by means of the study of the values needed in the explanatory variables for reaching such potential “optimal” values in the response variables. Overall, the main novelty of this post-econometric study is that the type of information obtained with multiobjective optimization approaches can be hardly reached just by the analysis of the problem from the econometric point of view.

Next, we describe in details how to apply the theoretical framework to combine econometrics and multiobjective optimization. It consists of two stages: the first one comprises the econometric analysis of the socio-economic phenomenon under study, while, in the second one, a multiobjective optimization problem is built and studied.

### 3.1. First stage

At the first stage, an econometric analysis of the socio-economic problem considered is performed to estimate the response variables based on the explanatory variables as in equation (2.1), giving raise to predictors  $\hat{y}^j(x_1, \dots, x_k)$  of the response variables ( $j = 1, \dots, m$ ). As previously said, these estimations can be obtained using different models, such as linear, polynomial, exponential, etc.

A detailed analysis of the regression model must be carried out in order to detect statistically significant correlations among the variables showing clear dependencies among them. Actually, to be able to apply our proposal, there must exist some degree of conflict among the response variables, in the sense that an improvement of some of them needs to imply a sacrifice in some of the others. It is important to detect that this conflicting nature exists among the response variables, since this will support the analysis of the problem considered from a multiobjective optimization point of view.

In addition, the possible correlations among the explanatory variables must be studied to identify possible dependencies between them, according to the data considered. As explained hereafter in the second stage, these dependencies will allow us to define the feasible variable values taking into account only realistic and meaningful values based on the dataset used.

### 3.2. Second stage

The second stage consists of the formulation of a multiobjective optimization problem based on the econometric analysis. To this aim, we define the decision variables, the constraints and the objective functions as follows.

#### 3.2.1. Decision variables

The explanatory variables  $x_1, \dots, x_k$  considered as key variables of interest in the econometric model (*i.e.* significant as regressors of the response variables) constitute the decision variables of the multiobjective optimization problem (we also denote them by  $x_1, \dots, x_k$  to simplify the notation). Depending on the situation studied, the explanatory variables can be continuous, integer or binary, so the decision variables will have the same nature.

Special care must be taken when binary categorical variables are included as regressors -to represent attributes- in the econometric model. In such a situation, a reference variable is normally introduced to control any possible scenario regarding the attribute considered and, usually, this reference variable is assumed to equal 1 if the rest of the binary categorical variables are 0. In the multiobjective optimization problem, the

reference variables associated to binary categorical variables are never considered as decision variables. Instead, the possible values that these binary categorical variables can take will be controlled by means of constraints, as explained hereafter.

### 3.2.2. Constraints

Next, the feasible set of the problem is built (through a set of constraints), taking into account both the meaning of the explanatory variables and the econometric analysis undertaken in the first stage.

On the one side, in the presence of binary categorical variables, some technical constraints may be required to control two types of situations. Firstly, it may be needed to ensure that some binary categorical variables do not simultaneously take the value 1. For example, if the variables with sub-indexes in the subset  $H \in \{1, \dots, k\}$  cannot equal 1 at the same time, a constraint as  $\sum_{i \in H} x_i = 1$  can be included into the multiobjective optimization problem. Secondly, it may be necessary to avoid the consideration of reference variables as decision variables, in case a reference variable is assigned to a group of binary categorical variables, as above-mentioned. In this case, instead of an equality constraint as the previous one, an inequality constraint as  $\sum_{i \in H} x_i \leq 1$  can be introduced into the problem. To sum up, if  $\text{TC}^E$  and  $\text{TC}^I$  are the total number of equality and inequality technical constraints needed, respectively, and  $\{H_1, \dots, H_{\text{TC}^E}\}$  and  $\{H_1, \dots, H_{\text{TC}^I}\}$  are the subsets of binary categorical variables of the econometric model whose values need to be controlled, the technical constraints that will be considered in the problem are:

$$g_h^E(x_1, \dots, x_k) = \sum_{i \in H_h} x_i - 1 = 0, \quad \text{for } h = 1, \dots, \text{TC}^E, \quad (3.1)$$

$$g_h^I(x_1, \dots, x_k) = \sum_{i \in H_h} x_i - 1 \leq 0, \quad \text{for } h = 1, \dots, \text{TC}^I. \quad (3.2)$$

On the other side, other constraints need to be introduced in order to define the feasible set in a realistic way, assuring that only meaningful and sufficiently realistic values of the decision variables are possible according to the observational data considered. Note that there may exist explanatory variables showing strong dependencies whose values could not be set independently in the problem. These variables need to be controlled by bound constraints, which could be built from the regression analysis by means of the correlation coefficients and the confidence intervals. Therefore, in the first stage, a correlation analysis must be carried out to select highly significant joint variances and to build up confidence intervals (for the probability  $p$  desired, which is usually at least 95%). For example, if we assume that the correlation coefficient calculated in the first stage among two variables,  $x_s$  and  $x_t$  with  $s, t \in 1, \dots, k$  and  $s \neq t$ , is statistically significant but not enough to be dropped one of them out<sup>2</sup>, then  $x_s$  can be approximated using  $x_t$  as follows:

$$x_s = \hat{a} \cdot x_t + \hat{b} \quad \text{with} \quad \hat{a} \in [\hat{a}^l, \hat{a}^u] \quad \text{and} \quad \hat{b} \in [\hat{b}^l, \hat{b}^u], \quad (3.3)$$

where  $[\hat{a}^l, \hat{a}^u]$  and  $[\hat{b}^l, \hat{b}^u]$  are the confidence intervals for the correlation coefficients  $\hat{a}$  and  $\hat{b}$ , respectively, at the confidence level  $p$ . Let us remark that linearity has been assumed among them for simplicity, although any other regression model can be used if desired. According to (3.3), the variable  $x_s$  can take only values between the following bounds:

$$\hat{a}^l \cdot x_t + \hat{b}^l \leq x_s \leq \hat{a}^u \cdot x_t + \hat{b}^u.$$

This provides us the following bound constraints, which assure that the dependence between  $x_s$  and  $x_t$  is considered in the model:

$$\begin{aligned} x_s - (\hat{a}^l \cdot x_t + \hat{b}^l) &\geq 0, \\ x_s - (\hat{a}^u \cdot x_t + \hat{b}^u) &\leq 0. \end{aligned}$$

<sup>2</sup>In other words, there is not strong multicollinearity.

In general, if we observe that the correlation coefficient among a variable  $x_s$  with respect to a subset of variables  $x_{t1}, \dots, x_{tv}$ , with  $s \notin \{t1, \dots, tv\}$ , shows a statistically significant dependence at the certain confidence level  $p$ , we can estimate  $x_s$  as a linear function of  $x_{t1}, \dots, x_{tv}$ :

$$x_s = \hat{a}_1 \cdot x_{t1} + \dots + \hat{a}_v \cdot x_{tv} + \hat{b} \quad \text{with} \quad \hat{a}_1 \in [\hat{a}_1^l, \hat{a}_1^u], \dots, \hat{a}_v \in [\hat{a}_v^l, \hat{a}_v^u], \\ \text{and} \quad \hat{b} \in [\hat{b}^l, \hat{b}^u],$$

where the confidence intervals  $[\hat{a}_1^l, \hat{a}_1^u]$ ,  $[\hat{a}_v^l, \hat{a}_v^u]$  and  $[\hat{b}^l, \hat{b}^u]$  of  $\hat{a}_1$ ,  $\hat{a}_v$  and  $\hat{b}$ , respectively, explain the sample data at the confidence level  $p$ . To allow only solutions whose values for  $x_s$  and  $x_{t1}, \dots, x_{tv}$  are not independent one from each other, assuring that they are within the confidence interval at the level  $p$ , the following two bound constraints will be introduced into the problem:

$$x_s - (\hat{a}_1^l \cdot x_{t1} + \dots + \hat{a}_v^l \cdot x_{tv} + \hat{b}^l) \geq 0, \\ x_s - (\hat{a}_1^u \cdot x_{t1} + \dots + \hat{a}_v^u \cdot x_{tv} + \hat{b}^u) \leq 0.$$

Similarly, other non-linear relationships existing among the variables may imply the introduction of other constraints into the multiobjective optimization problem. Then, in general, if we assume that a statistically significant non-linear dependency is observed between a variable  $x_s$  with respect to  $x_{t1}, \dots, x_{tv}$ , with  $s \notin \{t1, \dots, tv\}$ , as follows:

$$x_s = \bar{h}_s(x_{t1}, \dots, x_{tv}) + \hat{c}_s,$$

where  $h_s$  is a non-linear function of the variables  $x_{t1}, \dots, x_{tv}$  and the constant term  $\hat{c}_s$  is within a confidence interval  $[\hat{c}_s^l, \hat{c}_s^u]$  at a level  $p$ , we can formulate the following constraints to generalize the ones given for the linear case:

$$x_s - (\bar{h}_s(x_{t1}, \dots, x_{tv}) + \hat{c}_s^l) \geq 0, \\ x_s - (\bar{h}_s(x_{t1}, \dots, x_{tv}) + \hat{c}_s^u) \leq 0.$$

To group this type of constraints (linear and non-linear), let us denote by IC the total number of explanatory variables  $x_s$  that show a statistically significant dependence with respect to several explanatory variables  $x_{t1}, \dots, x_{tv}$ , with  $s \notin \{t1, \dots, tv\}$ . Then, the constraints built as above-explained can be formulated as follows, with  $s = 1, \dots, \text{IC}$ :

$$h_{2s-1}(x_1, \dots, x_k) = (\bar{h}_s(x_{t1}, \dots, x_{tv}) + \hat{c}_s^l) - x_s \leq 0, \quad (3.4)$$

$$h_{2s}(x_1, \dots, x_k) = x_s - (\bar{h}_s(x_{t1}, \dots, x_{tv}) + \hat{c}_s^u) \leq 0. \quad (3.5)$$

For example, in the above-described linear case, for  $s = 1, \dots, \text{IC}$ , the constraints (3.4) and (3.5) would be given by:

$$h_{2s-1}(x_1, \dots, x_k) = (\hat{a}_1^l \cdot x_{t1} + \dots + \hat{a}_v^l \cdot x_{tv} + \hat{b}^l) - x_s \leq 0, \\ h_{2s}(x_1, \dots, x_k) = x_s - (\hat{a}_1^u \cdot x_{t1} + \dots + \hat{a}_v^u \cdot x_{tv} + \hat{b}^u) \leq 0.$$

With all these constraints, we avoid solutions that do not fit the reality of the phenomenon studied according to the observational data available. As can be seen, the reliability of the feasible set obtained in this way depends on the rigour of the correlation analysis carried out in the first stage. Therefore, it is important to perform a detailed analysis of the dependency among the explanatory variables in the first stage, before formulating the multiobjective optimization problem in the second stage of the framework.

### 3.2.3. Objective functions

The purpose of the multiobjective optimization problem is to optimize all the response variables at the same time. Since the econometric study has allowed us to express the response variables as predictor functions of the explanatory variables, the objective functions of the problem, denoted by  $f_j(x_1, \dots, x_k)$  for  $j = 1, \dots, m$ , are formulated using the expected value  $\hat{y}_j$  of each response variable  $y_j$  obtained through the econometric model:

$$f_j(x_1, \dots, x_k) = \hat{y}_j(x_1, \dots, x_k), \quad \text{for } j = 1, \dots, m. \quad (3.6)$$

### 3.2.4. The multiobjective optimization model

Therefore, if we denote the vector of decision variables by  $\mathbf{x} = (x_1, \dots, x_k)^T$ , the multiobjective optimization problem built in the second stage can be defined as follows<sup>3</sup>, according to the objective functions given in (3.6) and the constraints in (3.1)–(3.5):

$$\begin{aligned} \max \quad & \{\hat{y}_1(\mathbf{x}), \hat{y}_2(\mathbf{x}), \dots, \hat{y}_m(\mathbf{x})\} \\ \text{subject to } & g_h^E(\mathbf{x}) = 0, & \text{for } h = 1, \dots, \text{TC}^E, \\ & g_h^I(\mathbf{x}) \leq 0, & \text{for } h = 1, \dots, \text{TC}^I, \\ & h_s(\mathbf{x}) \leq 0, & \text{for } s = 1, \dots, 2 \cdot \text{IC}, \\ & l_i \leq x_i \leq u_i, & \text{for } i = 1, \dots, k, \end{aligned} \quad (3.7)$$

where  $l_i$  and  $u_i$  are lower and upper bounds for each variable  $x_i$  ( $i = 1, \dots, k$ ), respectively, which can be set according to the observational data.

In particular, if all the objective functions are linear and all functions  $h_s(\mathbf{x})$  of the constraints (3.4) and (3.5) are also linear, for  $s = 1, \dots, 2 \cdot \text{IC}$ , problem (3.7) would be a linear mixed integer multiobjective optimization problem<sup>4</sup>. Furthermore, in the linear case, it is possible to go a little further to better reflect the reality of the data in our model. To formulate the objective functions, we can use the confidence intervals of the coefficients  $\hat{\alpha}^j, \hat{\beta}_1^j, \dots, \hat{\beta}_k^j$ , reflecting the desired percentage  $p$  of the observational data (usually, at most, 95%). In this case, the objective functions  $\hat{y}_j$  will have the following form, for  $j = 1, \dots, m$ :

$$\hat{y}^j(\mathbf{x}) = \hat{y}^j(x_1, \dots, x_k) = [\alpha_j^L, \alpha_j^U] + [\beta_{j1}^L, \beta_{j1}^U] \cdot x_1 + \dots + [\beta_{jk}^L, \beta_{jk}^U] \cdot x_k,$$

where  $[\beta_{ji}^L, \beta_{ji}^U]$  is the confidence interval for the estimated value of the coefficient  $\hat{\beta}_{ji}$ , for every  $j = 1, \dots, m$ ,  $i = 1, \dots, k$ , and  $[\alpha_j^L, \alpha_j^U]$  is the one for  $\hat{\alpha}^j$ , for every  $j = 1, \dots, m$ , at the confidence level  $p$  desired. Then, problem (3.7) will be an interval multiobjective optimization problem as follows:

$$\begin{aligned} \max \quad & \left\{ [\alpha_1^L, \alpha_1^U] + \sum_{i=1}^k [\beta_{1i}^L, \beta_{1i}^U] \cdot x_i, \dots, [\alpha_m^L, \alpha_m^U] + \sum_{i=1}^k [\beta_{mi}^L, \beta_{mi}^U] \cdot x_i \right\} \\ \text{subject to } & g_h^E(\mathbf{x}) = 0, & \text{for } h = 1, \dots, \text{TC}^E, \\ & g_h^I(\mathbf{x}) \leq 0, & \text{for } h = 1, \dots, \text{TC}^I, \\ & h_s(\mathbf{x}) \leq 0, & \text{for } s = 1, \dots, 2 \cdot \text{IC}, \\ & l_i \leq x_i \leq u_i, & \text{for } i = 1, \dots, k. \end{aligned}$$

<sup>3</sup>Without loss of generality, all the objectives (*i.e.* all the response variables) have been supposed to be maximized. In case any of them must be minimized, it can be transformed in the maximization form just by multiplying by “−1”.

<sup>4</sup>A multiobjective optimization problem is linear if the objective functions and the functions which define the constraints are all linear. Besides, it is said to be mixed integer if it has continuous and integer (which can be binary) decision variables.



#### 4. SOLVING THE ECONOMETRIC-MULTIOBJECTIVE OPTIMIZATION PROBLEM

There are plenty of techniques to solve the multiobjective optimization problem (3.7) formulated in the framework. Multiple Criteria Decision Making (MCDM) [9] and Evolutionary Multiobjective Optimization (EMO) [7] are two of the most active research fields in multiobjective optimization.

On the one hand, MCDM methods usually require preferential information from a decision maker to find a single solution, called the most preferred solution. Here, preferential information refers to information that the decision maker expresses about the conflicting objectives or information about the purpose of the study, such as *e.g.* marginal rates of substitution, surrogate values for trade-offs, selection of a solution among a set of solutions, classification of objective functions, and reference values or goals [23]. Depending on the moment when the preferences are considered into the solution process, MCDM methods are classified into *a priori*, *a posteriori* and interactive methods. Technically, many MCDM methods (such as reference point-based techniques, goal programming, compromise programming, generating techniques, etc.) scalarize the multiobjective optimization problem, which means that a single real-valued function is formulated taking into account the original objective functions and the preferential information given by the decision maker. The resulting scalarizing function is then minimized over the feasible set using an appropriate mathematical programming technique to find a Pareto optimal solution that fits the preferences considered as much as possible. For further details, see [17, 23].

On the other hand, EMO algorithms work with a population of solutions and attempt to find a good approximation of the entire Pareto optimal front, achieving convergence and diversity by applying operators that simulate the natural evolution of the species, such as selection, crossover and mutation. In practice, EMO algorithms are able to handle problems of different nature, *i.e.* they can easily handle non-convex, non-differentiable or discontinuous objective functions, with binary and integer-valued variables. Furthermore, the so-called preference-based EMO algorithms incorporate some preferential information into the evolutionary algorithm in order to guide the search for new solutions towards the subset of the Pareto optimal front that best suits these preferences. For more information, see [5, 7].

Focusing back at our contribution, depending on the kind of multiobjective optimization problem formulated in the second stage, and on the type of information that wants to be extracted in the study, the multiobjective optimization technique used to solve problem (3.7) must be selected carefully. For example, [27] applied goal programming to a policy problem. In [30], the multiobjective optimization problem was linear and two interactive methods based on local trade-offs were used for solving it. In [20], a linear mixed integer seven-objective optimization model was formulated and solved by means of a reference point-based technique to determine the profile of the most satisfied Spanish worker. In addition, they applied a combined goal programming reference point approach to identify policies that could increase workers satisfaction levels. The multiobjective optimization problems proposed in [18, 21] were also linear and mixed integer, and its solution was found using a reference point-based approach, whose robustness was analyzed by means of a sensitivity analysis. Goal programming was applied in [10] to find which policies could be carried out to increase insurers' results, using desirable targets for the objective functions. In [11, 12], multiobjective interval programming models were formulated and solved by multiobjective interval programming techniques to disentangle the extent to which the correlations found may be affected.

Some of the aforementioned works obtained linear and quadratic multiobjective optimization problems which could be easily solved by classical MCDM techniques. However, when the regression models are non-linear and non-quadratic, the resulting problem has not so ideal features and other techniques such as *e.g.* EMO algorithms may be needed to solve it. Additionally, in the case that confidence intervals are considered for the linear case, it is necessary to apply techniques of interval multiobjective programming [26]. Nevertheless, the methodology used has to be chosen according to the characteristics of the problem and taking into account the purpose of the study being performed.

Whatever the solving technique is applied, it is important to understand the meaning of the Pareto optimal solution/s obtained in the context of the studied phenomenon. According to the formulation given in Section 3, each Pareto optimal solution of problem (3.7) gives values for the decision variables (*i.e.* the explanatory

variables) that imply the simultaneous achievement of “optimal” values for the objective functions (*i.e.* the response variables). That is, by obtaining a Pareto optimal solution, optimal values for the response variables can be identified (based on to the data observed) and we can detect which explanatory variables have a great influence to reach such optimal values. Let us remark that, in the proposed framework, each Pareto optimal solution describes which combination of explanatory variable values must be promoted with policies and decisions if such “optimal” values of the response variables are desired to be achieved in the future.

#### 4.1. Post-optimization and sensitivity analysis

After solving the multiobjective optimization problem, a post-optimization analysis of the results obtained can lead to wider findings than those provided just with the econometric analysis. We can study how sensitive the “optimal” response variable values are to small changes in the explanatory variable values. This post-optimization analysis may be carried out in multiple ways.

On the one hand, it is not correct to assume that the dependencies observed in the data will remain unchanged in the future and the relaxation of some of the constraints may help to understand the scope for some flexibility in terms of achievable targets. To this aim, a sensitivity analysis of the constraints of the multiobjective optimization problem proposed must be performed. Note that some of the constraints of the problem may be binding<sup>5</sup> at the final solution, what means in practice that small variations in these constraints may greatly affect the results. Furthermore, if the binding constraints at the final solution are of the type (3.4) and (3.5), which are built from the regressions of the econometric analysis, the explanatory variables at the final solution are being forced to stay within the limits imposed by the observational data. Therefore, it is reasonable to allow certain violation degree of these constraints to study if better objective function values (*i.e.* response variable values) can be achieved if they are relaxed. Goal programming [29] is probably a suitable multiobjective technique to study the so-called soft constraints which enable penalized violations of some of the original constraints. This post-optimisation analysis is especially useful to know the impact of a change of the explanatory variables on the different response variables, enabling us to have a better image of the possible future situation if certain decisions are made.

Besides, a sensitivity analysis of the final solution can help to check its robustness regarding some of the factors influencing in the definition of the problem. For example, if a method based on reference points is used, the robustness can be analysed by studying how the final solution would vary with respect to the reference values and/or weights used. This analysis strengthens the methodology used since it investigates the consistency of the results obtained.

In addition, some further remarks must be taken into account. As above-described, the multiobjective optimization problem (3.7) has been formulated from the econometric analysis. Note that, in some situations, there may exist one or several explanatory variables that may not be susceptible of change because of their nature, such as *e.g.* age, education level of parents, and so on. Despite of their uncontrollable nature, these explanatory variables are considered as decision variables in the problem, and their values at the final solution must be understood as desirable or ideal values for them if we want to achieve the “best possible” response variable values.

Finally, it is important to clarify that the interaction with the decision maker in the framework proposed is not intended at just searching for a final solution to the problem (3.7). In comparison to the role of the analyst in econometrics -who is just in charge of analysing the data, here the decision maker interacts with the solution process by expressing her/his hopes and preferences with respect to the response variables (our objective functions). Apart from the optimal values obtained according to these preferences, this interaction aims at identifying which policies must be promoted or improved regarding the explanatory variables (decision variables) in order to achieve in the future such optimal values. Also, the final solution must be understood as a way to support and reinforce arguments obtained with the econometric analysis, such as expected values of

---

<sup>5</sup>A constraint is *binding* at a solution if the inequality is met with equality at this solution. Otherwise, it is said to be *non-binding*.

the explanatory variables under certain scenarios, which may help to define meaningful and consistent policies for the future.

## 5. ILLUSTRATIVE EXAMPLE

In this section, we demonstrate an application of the methodology proposed. The phenomenon studied in this example is related to economics of education and it is briefly introduced in what follows.

### 5.1. Introduction about the application problem

Recently, the study of students' well-being is receiving a great deal of attention because of its causal relation with the academic achievement of young people [22, 31]. High levels of well-being are related to students with positive and fulfilling life-experiences, while low levels of it may imply just the opposite. Furthermore, education is a key factor in the economic growth of any country [3], so improving the students' well-being can be beneficial not only for the students themselves but also for the economy of the country.

To improve the students' academic results by increasing their levels of well-being, on the one hand, it is important to decide which aspects of the students' life enable us to quantify their levels of well-being [4]. Note that the concept of well-being is intrinsically multidimensional, consisting of several cognitive, psychological, social, and physical characteristics [15, 24]. Since 2015, the PISA report<sup>6</sup> provides for some countries a complete set of well-being indicators, which are built according to the students' answers to an additional questionnaire related to their well-being (for further information about the building process of these indexes, see [25]). From the available ones, we employ the following four indexes in our study: level of anxiety at school, student motivation, sense of belonging at school, and bullying<sup>7</sup>. Several studies support the analysis of the well-being by means of these indicators (*e.g.* see [6, 13, 16, 28]).

In relation to these four well-being indicators, the student with the "ideal" well-being would be the one who achieves the best possible values for the four of them. Ideally, the anxiety and the bullying indicators must achieve the lowest levels, while the motivation and the sense of belonging at school indexes are desired to reach the highest values. Nevertheless, as it is shown later, achieving reasonably good levels of all the indicators at the same time is not a straightforward matter, because they are in conflict (according to the analysis described next). This conflicting nature makes it impossible to have an "ideal" student well-being who reaches the desirable optimum levels of the four well-being indicators simultaneously. Therefore, it would be worth to have information about the sacrifices to pay in some of the indexes to reach an improvement at some of the others (trade-offs).

Besides, it is also important to identify the significant variables of the teaching-learning environment that would allow achieving the best possible levels of the indicators. That is, we need to know how the best indicator values could be obtained simultaneously, and which these values are, in order to gain some knowledge about the optimal educational context that must be promoted to improve the students' well-being. The factors used as key educational variables are related to socio-demographic features of the students, their abilities with the information and communication technologies (ICTs) and internet, and learning hours, among others. In our study, we have used personal and academic characteristics data of 15-year-old Spanish students from the PISA database corresponding to 2015.

Thus, the main purpose is to study and analyze which variables regarding the educational context are relevant, and which values should they get, in order to achieve, at the same time, acceptable levels of the indicators considered to quantify the well-being of students. By determining the profile of the student associated with the best indicator values, we can reach conclusions and we may identify policies that must be furthered by educational policy makers to improve the youngsters' well-being in the future, as a way to increase their academic performance.

<sup>6</sup>PISA is the OECD's Programme for International Student Assessment, which measures 15-year-old students' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges (<http://www.oecd.org/pisa/>).

<sup>7</sup>We have replicated these indexes through factor analysis to obtain them for an extended sample of Spanish students, since PISA calculates them just for a reduced sample [25].

## 5.2. Application of the theoretical framework

Our theoretical framework can be applied to reach the desired information about the four well-being indicators (*i.e.* which optimum values they can attain simultaneously, and how these values can be obtained by means of the considered educational variables).

In the first stage of the framework, the econometric analysis will identify the relationships among the variables considered, and if they can be used to regress the well-being indicators through an econometric model. Actually, with a careful analysis of the regressions, we will also determine to which extent the four well-being indicators are in conflict. This will confirm that their optimal values are difficult to achieve simultaneously, and will justify the analysis of this socio-economic problem from a multiobjective optimization point of view.

Next, a multiobjective optimization problem will be formulated in the second stage, at which the predictors of the well-being indicators will define the objective functions to be optimized. Next, the application of multiobjective optimization techniques will allow us to gain some knowledge about the trade-offs existing between the four indicators. Based on this information, policy makers will be able to understand how the improvement of one indicator (anxiety, motivation, sense of belonging or bullying) may affect the performance of the others, enabling them to anticipate the impacts of the possible education policies on the students' well-being.

### 5.2.1. First stage: Econometric analysis

Initially, let us statistically analyze the empirical information considered in this study, which comes from the database PISA 2015. In total, 32 330 15-year-old students from Spain participated in the assessment. Our analysis is focused on non-repeater students, and students enrolled in public and semi-private schools (private schools are not representative). Considering the missing data, the sample used was formed by 17 128 Spanish students.

All the information about the educational variables considered can be found in Table 1 and their descriptive statistics are given in Table 2. The data indicates that most of the students (69.5%) of our sample belongs to a public school, while only 30.5% comes from semi-private ones. The proportion of girls and boys and the date of birth are balanced. Indeed, on average, Spanish students use internet three hours per day outside school, and started using digital devices and internet at the ages of seven and eight, respectively. Furthermore, they spend about four hours per week studying math, and three hours per week reading at home.

Table 2 also shows the mean and standard deviation attained by the four well-being indicators, according to our data<sup>8</sup>. Remember that these indicators are composite indexes that synthesize students' answers to different questions, so they do not have values with an understandable meaning. Nevertheless, what it is the most important to interpret their meaning is that the anxiety and the bullying indexes should achieve the lowest possible levels, while the motivation and the sense of belonging are desired to be as higher as possible.

As described in Section 3.1, the first step consists of an econometric analysis of the data. In order to observe the correlation between students' well-being indicators and the different educational variables, we have estimated regression models – by ordinary least squares (OLS). To this aim, we consider as response variable the four indicators – anxiety index ( $y_1$ ), motivation index ( $y_2$ ), sense of belonging index ( $y_3$ ) and bullying index ( $y_4$ ) –, while the explanatory variables are the educational variables given in Table 1 – denoted as  $x_i$ , for  $i = 1, \dots, 11$ .

Then, if  $r$  represents the order of the students ( $r = 1, \dots, n$ , with  $n = 17\,128$ ), the regression model obtained by OLS is defined as follows:

$$\hat{y}_j(r) = \hat{\alpha}_j + \hat{\beta}_1^j \cdot x_1(r) + \hat{\beta}_2^j \cdot x_2(r) + \dots + \hat{\beta}_{11}^j \cdot x_{11}(r) + \epsilon_j(r), \quad (5.1)$$

where  $\hat{y}_j(r)$  estimates the response variable  $y_j(r)$ , for  $j = 1, 2, 3, 4$ ,  $x_1(r), \dots, x_{11}(r)$  are the set of explanatory variables for each student  $r$ ,  $\epsilon_j(r)$  is a random disturbance for the student  $r$ ,  $\hat{\beta}^j = (\hat{\beta}_1^j, \dots, \hat{\beta}_{11}^j)^T$  is the vector of slope coefficients, and  $\hat{\alpha}_j$  a fixed but unknown population intercept (constant). Table 3 shows the estimated

<sup>8</sup>Note that these indicators have been obtained through factor analysis and thus, they have a mean of 0 and a standard deviation equal to 1, approximately.

TABLE 1. Educational variables under scrutiny.

Variable	Notation	Description	Values
Socio-economic index	$x_1$	Index of economic, social and cultural status	$[-4.284, 3.271]$
Semi-private school	$x_2$	Type of school (Semi-private 1 – public 0)	0 or 1
Period of birth:		Trimester of birth (reference group: the first one)	
Second period	$x_3$	Second period (April, May and June)	0 or 1
Third period	$x_4$	Third period (July, August and September)	0 or 1
Fourth period	$x_5$	Fourth period (October, November and December)	0 or 1
Girls	$x_6$	Girls	0 or 1
Hours of internet outside school	$x_7$	Amount of time (hours) students spend connected to the Internet during a typical week	$[0, 7]$
Age at use of digital devices	$x_8$	How old were you when you first used a digital device?	$[4.5, 14.5]$
Age at use of internet	$x_9$	How old were you when you first accessed the Internet?	$[4.5, 14.5]$
Hours of math	$x_{10}$	This school year, approximately how many hours per week do you spend learning in addition to your required school schedule in math?	$[0, 30]$
Hours of reading	$x_{11}$	This school year, approximately how many hours per week do you spend learning in addition to your required school schedule in reading?	$[0, 30]$

coefficients of the significant variables for the four indicators, reporting also the standard deviations (in parentheses) and the significance levels for the estimated coefficients (respectively, the super-index \*\*\* means that the estimation is significant at 1%, \*\* significant at 5%, and \* significant at 10%).

Regarding the results of the estimations, it can be seen that the socio-economic index has a negative effect in the anxiety index, supporting the idea that students with a higher socio-economic status tend to show lower levels of anxiety than those coming from more modest families. This variable affects positively to the motivation and sense of belonging indexes, both achieving higher levels as the socio-economic level increases. Concerning the fact of attending to a public or a semi-private school, the estimators reveal that students in public schools have a lower feeling of belonging and experience less bullying than those in semi-private schools. This shows a conflicting relation among the sense of belonging and the bullying indexes, at least regarding the type of school the students are enrolled: attending to a semi-private school implies an improvement for the sense of belonging index but a sacrifice regarding the bullying, while those students in public schools are expected to enhance the bullying index at the expense of a decrease of their sense of belonging.

Concerning the gender, our findings are consistent with other studies (see *e.g.* [19]) since the anxiety seems to achieve higher levels for girls than for boys. Usually, girls seem to feel more responsible than boys do and, therefore, studying and doing well causes them more anxiety. On the contrary, the motivation is revealed to be higher for boys than for girls, in line with *e.g.* [1]. Nevertheless, our results indicate that boys are more likely to be victims of bullying than girls. As a conclusion, we can state that girls suffer more anxiety and face less

TABLE 2. Descriptive statistic.

		Type	Mean	Standard deviation
<i>Well-being indexes</i>				
Anxiety index	$y_1$	Continuous	0.036	0.802
Motivation index	$y_2$	Continuous	0.150	0.828
Sense of belonging index	$y_3$	Continuous	0.005	0.872
Bullying index	$y_4$	Continuous	-0.032	0.815
<i>Educational variables</i>				
Socio-economic index	$x_1$	Continuous	-0.289	1.121
Semi-private school	$x_2$	Binary	0.305	0.461
Period of birth:				
Second period	$x_3$	Binary	0.248	0.432
Third period	$x_4$	Binary	0.261	0.439
Fourth period	$x_5$	Binary	0.240	0.429
Girls	$x_6$	Binary	0.533	0.499
Hours of internet outside school	$x_7$	Continuous	3.390	2.156
Age at use of digital devices	$x_8$	Continuous	7.349	2.613
Age at use of internet	$x_9$	Continuous	8.624	2.454
Hours of math	$x_{10}$	Continuous	4.542	3.665
Hours of reading	$x_{11}$	Continuous	3.666	3.328

bullying than boys, but they are less motivated than boys. Again, this indicates that the anxiety, the motivation and the bullying are conflicting, since when any of them improves its value, some of the others get worse values, at least within boys and girls.

Regarding the use of ICTs, the number of hours dedicated to internet at home is positively correlated with the anxiety, the sense of belonging and the bullying indexes, and seems to have a negative effect in the motivation. This reveals that using internet outside school during long periods: increases the levels of anxiety [16]; improves the belonging feeling to the school [2]; is associated with higher levels of bullying; and, finally, affects negatively the academic motivation of students [8]. All of this shows a clear conflict among the sense of belonging indicator and the rest of indicators, since this indicator gets better values when the other three achieve worse levels. In addition, the results suggest that, when students start using digital devices and internet at higher ages, the levels of anxiety tend to be higher. Indeed, students seem to be less (respectively, more) motivated if their first contact with a digital device happened when they were older (respectively, younger).

Finally, our estimations enable us to conclude that spending more time studying math and reading at home is related with higher levels of anxiety, which may be a negative effect of the responsibility these students impose themselves for reaching high scores. However, dedicating more time to study math is positively associated with the motivation and the sense of belonging, while students who spend more time reading are related to more bullying. Therefore, in relation to the time dedicated to studying math, a conflict is observed between the anxiety level and the motivation and sense of belonging indexes, given that the former is higher (respectively, lower) when the levels of the two latter increase (respectively, decrease).

As shown, our correlation analysis has corroborated that certain degrees of conflict do exist among the four well-being indexes. Remember that, ideally, the anxiety and bullying are desired to achieve the lowest levels (*i.e.* to be minimized), while the motivation and sense of belonging indexes should be as higher as possible (*i.e.* should be maximized). This conflicting nature justifies the application of a multiobjective optimization approach such as the one at the second stage of our framework, to extract further conclusions about the simultaneous achievement of optimal compromise values of the four indicators.

In general, when applying our framework to other scenarios, this is the most important information that needs to be found from the econometric analysis in the first stage. Note that having a conflict among the response

TABLE 3. OLS estimated coefficients for the students' well-being indexes (normalized).

Explanatory variables	Response variables			
	Anxiety ( $\hat{\beta}_i^1$ )	Motivation ( $\hat{\beta}_i^2$ )	Sense of belonging ( $\hat{\beta}_i^3$ )	Bullying ( $\hat{\beta}_i^4$ )
Socio-economic index $x_1$	-0.063*** (0.009)	0.104*** (0.008)	0.017** (0.008)	0.004 (0.010)
Semi-private school $x_2$	-0.015 (0.023)	0.031 (0.026)	0.068*** (0.026)	0.085*** (0.026)
Period of birth:				
Second period $x_3$	0.002 (0.022)	-0.033 (0.025)	0.050* (0.026)	-0.028 (0.023)
Third period $x_4$	0.028 (0.023)	-0.030 (0.021)	0.105*** (0.026)	-0.018 (0.026)
Fourth period $x_5$	0.043* (0.025)	-0.034 (0.022)	0.056** (0.028)	-0.004 (0.023)
Girls $x_6$	0.327*** (0.017)	-0.082*** (0.016)	-0.024 (0.021)	-0.137*** (0.019)
Hours of internet outside school $x_7$	0.006* (0.004)	-0.012*** (0.003)	0.011** (0.004)	0.014*** (0.004)
Age at use of digital devices $x_8$	0.017*** (0.004)	-0.015*** (0.004)	-0.000 (0.005)	0.006 (0.006)
Age at use of internet $x_9$	0.007* (0.004)	-0.004 (0.004)	-0.005 (0.005)	-0.001 (0.006)
Hours of math $x_{10}$	0.021*** (0.004)	0.010*** (0.003)	0.009** (0.004)	-0.003 (0.003)
Hours of reading $x_{11}$	0.009** (0.004)	-0.005 (0.003)	-0.007 (0.004)	0.010** (0.005)
Constant $\alpha_j$	-0.504	0.398	-0.061	0.072
Observations	17 128	17 128	17 128	17 128
R-squared	0.0825	0.0334	0.0061	0.0111

variables that are estimated must be guaranteed to study the problem from a multiobjective optimization perspective at the second stage.

5.2.2. Second stage: building a multiobjective optimization problem

Now, we are able to build the multiobjective optimization problem from the econometric analysis, as described in Section 3.2. In this problem, the decision variables are the explanatory variables of the econometric model (i.e. the variables in Tab. 1). To have a model adjusted as much as possible to the reality observed, the feasible set of possible solutions must be built taking into account the meaning of the variables and the findings of the econometric analysis, as said in Section 3.2.2. In this case, we need to add just one technical constraint in relation to the variables controlling the period of birth (i.e.  $x_3, x_4, x_5$ ) to assure that only one of these three binary variables achieves the value 1:

$$g_1^I(x_1, \dots, x_{11}) = x_3 + x_4 + x_5 - 1 \leq 0.$$

In addition, the problem must also consider some constraints formulated according to the strong dependencies observed in a linear regression analysis performed for all the possible combinations among the variables. As explained in Section 3.2.2, two bound constraints as the ones given in equations (3.4) and (3.5) are built with respect to the variables showing a significant correlation, using 95% confidence intervals. Overall, 10 constraints were obtained as (3.4) and (3.5) using the significant dependencies between the variables. The bounds of the

TABLE 4. Bound constraints based on dependency among the explanatory variables.

Linear dependency	Constraints	$\hat{a}_1^l$	$\hat{a}_2^l$	$\hat{b}^l$	$\hat{a}_1^u$	$\hat{a}_2^u$	$b^u$
$x_7 = \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2 + \hat{b}$	C1	-0.241	0.249	3.064	-0.132	0.452	3.232
	C2						
$x_8 = \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2 + \hat{b}$	C3	-0.353	0.498	6.840	-0.245	0.758	7.064
	C4						
$x_9 = \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2 + \hat{b}$	C5	-0.394	0.304	8.233	-0.301	0.528	8.372
	C6						
$x_{10} = \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2 + \hat{b}$	C7	0	0.230	4.193	0	0.561	4.473
	C8						
$x_{11} = \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2 + \hat{b}$	C9	-0.152	0.028	3.428	0.017	0.335	3.662
	C10						

TABLE 5. Lower and upper bounds of the explanatory variables.

Lower bound ( $l_i$ )	Variable	Upper bound ( $u_i$ )
-4.728	$x_1$	3.271
0	$x_2$	1
0	$x_3$	1
0	$x_4$	1
0	$x_5$	1
0	$x_6$	1
0	$x_7$	7
4.5	$x_8$	14.5
4.5	$x_9$	14.5
0	$x_{10}$	30
0	$x_{11}$	30

confidence intervals needed to build them are given in Table 4, where we also indicate the variables involved in each case. For simplicity, these constraints have been named as (C1) to (C10).

For the sake of clarity, an example is explained next to understand the information given in Table 4. In our analysis, the variable  $x_7$  showed a significant dependence with respect to  $x_1$  and  $x_2$ . A linear relation between them was observed as  $x_7 = \hat{a}_1 \cdot x_1 + \hat{a}_2 \cdot x_2 + \hat{b}$ , with the following confident intervals of the coefficients (at 95%):  $\hat{a}_1 \in [\hat{a}_1^l, \hat{a}_1^u] = [-0.241, -0.132]$ ,  $\hat{a}_2 \in [\hat{a}_2^l, \hat{a}_2^u] = [0.249, 0.452]$  and  $\hat{b} \in [\hat{b}^l, \hat{b}^u] = [3.064, 3.232]$ . According to Section 3.2.2, this implies that two bound constraints as (3.4) and (3.5) are built as follows (which correspond to the constraints (C1) and (C2) in Tab. 4):

$$\begin{aligned}
 h_1(x_1, \dots, x_{11}) &= (\hat{a}_1^l \cdot x_1 + \hat{a}_2^l \cdot x_2 + \hat{b}^l) - x_7 \leq 0, \\
 h_2(x_1, \dots, x_{11}) &= x_7 - (\hat{a}_1^u \cdot x_1 + \hat{a}_2^u \cdot x_2 + \hat{b}^u) \leq 0. \Rightarrow \\
 &\Rightarrow 0.241 \cdot x_1 + 0.249 \cdot x_2 + 3.064 - x_7 \leq 0, \\
 &\quad x_7 + 0.132 \cdot x_1 - 0.452 \cdot x_2 - 3.232 \leq 0.
 \end{aligned}$$

All the information required to build these two constraints (the variables involved and the bounds of the confidence intervals) is summarized in the first two lines of Table 4. Similarly, the constraints (C3)–(C10) are defined with the rest of information provided in Table 4. Finally, we have used the minimum and maximum values attained by the variables according to our data (see Tab. 1) to define the lower and upper bounds of the variables provided in Table 5.



TABLE 6. Pay-off matrix for the multiobjective optimization problem (5.2).

		Anxiety (EI <sub>1</sub> )	Motivation (EI <sub>2</sub> )	Sense of belonging (EI <sub>3</sub> )	Bullying (EI <sub>4</sub> )
Optimal solution of EI <sub>1</sub>	$\mathbf{x}_1^*$	<b>-0.437</b>	0.668	0.057	0.133
Optimal solution of EI <sub>2</sub>	$\mathbf{x}_2^*$	-0.431	<b>0.670</b>	0.060	0.133
Optimal solution of EI <sub>3</sub>	$\mathbf{x}_3^*$	-0.402	0.657	<b>0.249</b>	0.232
Optimal solution of EI <sub>4</sub>	$\mathbf{x}_4^*$	-0.110	0.586	0.057	<b>-0.004</b>

Finally, the objective functions of the model are the response variables (*i.e.* the well-being indicators  $y_j$ , with  $j = 1, \dots, 4$ ): anxiety (to be minimized), motivation (to be maximized), sense of belonging (to be maximized) and bullying (to be minimized). To formulate these objective functions, we use the regressions  $\hat{y}_j$  obtained in the first stage that approximate these indexes as functions of the explanatory variables  $x_i$  (see Eq. (5.1)). Therefore, using the coefficients  $\hat{\beta}_i^j$  and  $\hat{\alpha}^j$  given in Table 3 ( $j = 1, \dots, 4, i = 1, \dots, 11$ ), the four objective functions are given by<sup>9</sup>:

$$EI_j(\mathbf{x}) = EI_j(x_1, \dots, x_{11}) = \hat{y}_j(x_1, \dots, x_{11}) = (\hat{\beta}^j)^T \mathbf{x} + \hat{\alpha}^j, \quad \text{for } j = 1, 2, 3, 4,$$

where  $(\hat{\beta}^j)^T = (\hat{\beta}_1^j, \dots, \hat{\beta}_{11}^j)^T$ ,  $\mathbf{x} = (x_1, \dots, x_{11})^T$  and  $EI_j(\mathbf{x})$  denotes the estimated value of the indicator  $y_j$  at solution  $\mathbf{x}$ .

Based on all this information, the multiobjective optimization problem to be solved is the following one:

$$\begin{aligned} & \min && \{EI_1(\mathbf{x}), EI_4(\mathbf{x})\} \\ & \max && \{EI_2(\mathbf{x}), EI_3(\mathbf{x})\} \\ & \text{subject to} && g_1^I(\mathbf{x}) \leq 0, \\ & && (C1) - (C10), \\ & && l_i \leq x_i \leq u_i, && \text{for } i = 1, \dots, 11. \end{aligned} \tag{5.2}$$

### 5.2.3. Solutions to the multiobjective optimization model

To have an initial idea about the trade-offs existing among the indicators, we have calculated their ideal values by individually optimizing each objective function over the feasible set, and this is the ideal vector obtained:

$$z^* = (-0.437, 0.670, 0.249, -0.004)^T.$$

The pay-off matrix shown in Table 6 displays the values of the four objective functions in each of their individual optima (for example, row 1 corresponds to the objective function values at the optimal solution for the anxiety index, row 2 shows the objective function values at the optimal solution for the motivation index, and so on). This matrix also demonstrates that there exists a conflict between each pair of functions, meaning that the Pareto optimal set is not limited to one solution. This also supports the analysis of the problem from a multiobjective optimization point of view. Indeed, observe that the values attained at  $\mathbf{x}_1^*$ ,  $\mathbf{x}_2^*$  and  $\mathbf{x}_3^*$  are quite far from the ideal value of the bullying indicator ( $-0.004$ ). The same happens if we compare the sense of belonging index values at  $\mathbf{x}_1^*$ ,  $\mathbf{x}_2^*$  and  $\mathbf{x}_4^*$  with its ideal value (0.249).

Let us recall that the aim of formulating and solving this multiobjective optimization problem is to find a compromise solution enabling to have balanced optimal values in the four well-being indicators, as a way to detect the educational context (*i.e.* the decision variable values) that would allow to achieve such “optimal”

<sup>9</sup>Only the significant coefficients  $\hat{\beta}_i^j$  -in any of the indicators, at 1%, 5%, 10%- were included in the final formulation of the objective functions.

situation. To this end, we have applied a reference point-based approach to solve problem (5.2). In this approach, the achievement scalarizing function proposed in [32] is minimized over the feasible set of solutions, in order to find the closest Pareto optimal solution to a reference point composed of desirable reference values for the objective functions. Mathematically, it is assured that any Pareto optimal solution of the original multiobjective optimization problem can be found by minimizing this function over the feasible set, using the ideal vector as reference point and modifying the weight vector in the whole weight vector space [23]. Actually, it is also demonstrated that any Pareto optimal solution can be found when fixing the weight vector and varying the reference point [23]. We would like to clarify that we have chosen a reference point-based approach just to show how the problem obtained can be explicitly solved, but the solving technique must be selected according to the problem obtained and to the type of findings that want to be reached about the situation under scrutiny.

We have considered the ideal values as desirable potential reference levels (denoted by  $q_j$ , for  $j = 1, \dots, 4$ ) of the four well-being indexes:

$$q_1 = -0.437, \quad q_2 = 0.670, \quad q_3 = 0.249, \quad q_4 = -0.004.$$

When applying the reference point-based approach, we have solved the differentiable formulation of the function proposed in [32], which is the following one for our problem (5.2):

$$\begin{aligned} \min \quad & \alpha + \rho((\hat{\beta}^1)^T \mathbf{x} + \hat{\alpha}^1 - q_1) + (q_2 - (\hat{\beta}^2)^T \mathbf{x} - \hat{\alpha}^2) \\ & + \rho((q_3 - (\hat{\beta}^3)^T \mathbf{x} - \hat{\alpha}^3) + ((\hat{\beta}^4)^T \mathbf{x} + \hat{\alpha}^4 - q_4)) \\ \text{subject to} \quad & (\hat{\beta}^j)^T \mathbf{x} + \hat{\alpha}^j \leq \alpha, & \text{for } j = 1, 4, \\ & q_j - (\hat{\beta}^j)^T \mathbf{x} - \hat{\alpha}^j \leq \alpha, & \text{for } j = 2, 3, \\ & g_1^I(\mathbf{x}) \leq 0, \\ & (C1) - (C10), \\ & l_i \leq x_i \leq u_i, & \text{for } i = 1, \dots, 11. \end{aligned} \quad (5.3)$$

In this formulation,  $\rho > 0$  is a so-called augmentation coefficient that assures the Pareto optimality of the solutions generated (we used  $\rho = 0.001$ ). All the criteria are equally weighted, meaning that it is implicitly assumed that the achievement of all the reference values have the same importance for the decision maker. The Pareto optimal solution obtained (values of the decision variables and of the objective functions) when solving problem (5.3) can be seen in Table 7.

Regarding the well-being indicator values at this solution, we can observe that the anxiety has achieved a value ( $-0.437$ ) very close to its ideal level -note that it is not exactly the same value since the sense of belonging at this solution gets a different value to the one at solution  $\mathbf{x}_1^*$  (see Tab. 6). Mathematically, this means that the Pareto optimal solution which minimizes the distance to the ideal vector has been reached, and this solution is very close to  $\mathbf{x}_1^*$  (the individual optimal solution for the anxiety index). In addition, the motivation index gets a value ( $0.668$ ) near to its ideal level. Nonetheless, the other two indicators (sense of belonging and bullying) are further from their ideal values ( $0.162$  and  $0.133$ , respectively). Then, we can conclude that, if we desire to simultaneously reach values for the four indicators as close as possible to their ideal levels, the best results are attained when the anxiety level is minimum, at the expense of a high sacrifice in the sense of belonging and bullying.

Focusing on the values obtained for the decision variables, we can extract several conclusions regarding the profile of the student getting the best possible indicator values. This solution indicates that the student associated with this Pareto optimal solution is a boy, born in the third period of the year, who is enrolled in a public school. His socio-economic status is high (actually, it has the highest possible socio-economic index according to our database). This student started using digital devices and internet at the ages of five and seven, respectively. On average, he spends two hours using internet outside the school, around four hours studying

TABLE 7. Solution to problem (5.3) for the multiobjective optimization problem (5.2).

Decision variables		Values
Socio-economic index	$x_1$	3.271
Semi-private schools	$x_2$	0.000
Period of birth:		
Second period	$x_3$	0.000
Third period	$x_4$	1.000
Fourth period	$x_5$	0.000
Girls	$x_6$	0.000
Hours of internet outside school	$x_7$	2.274
Age at use of digital devices	$x_8$	5.685
Age at use of internet	$x_9$	6.943
Hours of math	$x_{10}$	4.193
Hours of reading	$x_{11}$	2.930
Objective functions		
Anxiety index	EI <sub>1</sub>	-0.437
Motivation index	EI <sub>2</sub>	0.668
Sense of belonging index	EI <sub>3</sub>	0.162
Bullying index	EI <sub>4</sub>	0.133

math, and three hours reading at home. If we compare the values for the variables  $x_7, \dots, x_{11}$  with their mean values (see Tab. 2), we observe that the “optimal” student invest less time using internet, studying math and reading than the average, and the starting ages for using digital devices and internet are both earlier than the average (according to our sample).

This information shows that, if we want to decrease the anxiety and the motivation among students as much as possible, policies must be devoted to decrease the time spent at home for doing homework, which will decrease the stress levels of the students. In addition, the use of internet and digital devices have been revealed to have an impact to get optimal levels for these two indicators, so it must be promoted that students get used with these competences at earlier ages. Observe that this conclusion fits with the fact that the digitalization of many aspects of the current society makes all people be as much formed and updated as possible with the use of ICTs. If students are not sufficiently qualified in this regard, they may suffer from higher levels of anxiety and their motivation may decrease due to a lack of self-esteem. A remedy for this may be to get in contact with ICTs at earlier ages in the future, as we have concluded. Nevertheless, achieving this optimal scenario for the anxiety and the motivation would require a sacrifice, since the students may have a lower feeling of belonging to the school, and bullying is far from its ideal level. Thus, policies should be formulated with caution, taking into account this to prevent that these two aspects of the students’ well-being may be deteriorated in excess.

## 6. CONCLUSION

A mixed econometric-multiobjective optimization framework as the one proposed in this paper can be very useful for researchers and/or decision makers to study many socio-economic or management problems. The main contribution is that it is possible to identify values for the explanatory variables, fitting the observational data considered, that allow the simultaneous achievement of optimal levels for the response variables considered.

The proposed methodology is especially suitable when certain conflict is observed among the response variables considered. Based on the regressions obtained in the econometric analysis, and making use of the correlations found among the variables (first stage), a multiobjective optimization problem is built, whose solutions can provide interesting conclusions about the phenomenon studied (second stage).

With the resulting information, decision makers can define policies to improve the explanatory variables as much as possible in order to simultaneously achieve optimal levels for the response variables. They may also investigate how the dependencies existing between the explanatory variables can be altered in the future in order to reach satisfactory optimal values of the response variables. Actually, we can foresee the impact at some of the response variables that may cause the achievement of desirable optimal levels for some other response variables. This type of information is impossible to obtain just by an econometric analysis of the problem, so the framework proposed here constitutes a very relevant research contribution.

Finally, an illustrative example of economy of education has shown the functionality and potential of this methodology. As shown, the application of our proposal shed some light for the development of educational policies to promote the students' well-being, allowing to anticipate to the practical effects of improving the well-being indicators considered. We believe that the possibilities of implementing this methodological framework to other socio-economic contexts are immense and deserve future research work.

*Acknowledgements.* This work was supported by the Spanish Ministry of Science and Innovation (project PID2020-115429GB-I00), by the Andalusian Regional Ministry of Economy, Knowledge, Business and University (PAI group SEJ-532 and project UMA18-FEDERJA-024, also supported by FEDER funding), and by the University of Malaga (project B1-2020.18). Sandra Gonzalez-Gallardo is recipient of a research contract within "Ayudas para la Recualificación del Sistema Universitario Español, Modalidad Margarita Salas", financiado por la Unión Europea – NextGenerationEU.

## REFERENCES

- [1] M. Akib, H. Haryanto, I. Iskandar and A.A. Patak, Investigating the motivation, participation, and achievement of students. *Int. J. Humanit. Innovation* **1** (2018) 78–87.
- [2] S. Bannon, T. McGlynn, K. Mckenzie and E. Quayle, The positive role of internet use for young people with additional support needs: Identity and connectedness. *Comput. Human Behav.* **53** (2015) 504–514.
- [3] R.J. Barro, Human capital and growth. *Am. Econ. Rev.* **91** (2001) 12–17.
- [4] F. Borgonovi and J. Pál, A framework for the analysis of student well-being in the PISA 2015 study. *OECD Edu. Work. Papers* (2016).
- [5] J. Branke, K. Deb, K. Miettinen and R. Slowiński, editors. *Multiobjective Optimization – Interactive and Evolutionary Approaches*. Springer, Berlin, Heidelberg (2008).
- [6] M.M. Chiu, B.W. Chow, C. McBride and S.T. Mol, Students' sense of belonging at school in 41 countries: Cross-cultural variability. *J. Cross-Cult. Psychol.* **47** (2016) 175–196.
- [7] C.A.C. Coello, G.B. Lamont and D.A.V. Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd edition. Springer, New York (2007).
- [8] Y. Demir and M. Kutlu, Relationships among internet addiction, academic motivation, academic procrastination and school attachment in adolescents. *Int. Online J. Edu. Sci.* **10** (2018) 315–322.
- [9] M. Ehrgott and D. Tenfelde-Podehl, Computation of ideal and nadir values and implications for their use in MCDM methods. *Eur. J. Oper. Res.* **151** (2003) 119–139.
- [10] A.I. González-Fernández, M. Rubio-Misas and F. Ruiz, Goal programming to evaluate the profile of the most profitable insurers: An application to the Spanish insurance industry. *Int. Trans. Oper. Res.* **27** (2020) 2976–3006.
- [11] C.O. Henriques, M. Luque, O.D. Marcenaro-Gutierrez and L.A. Lopez-Agudo, A multiobjective interval programming model to explore the trade-offs among different aspects of job satisfaction under different scenarios. *Socio-Econ. Plan. Sci.* **66** (2019) 35–46.
- [12] C.O. Henriques, L.A. Lopez-Agudo, O.D. Marcenaro-Gutierrez and M. Luque, Reaching compromises in workers' life satisfaction: A multiobjective interval programming approach. *J. Happiness Stud.* **22** (2021) 207–239.
- [13] A.L. Jogi, E. Kikas, M.K. Lerkkanen and K. Mägi, Cross-lagged relations between math-related interest, performance goals and skills in groups of children with different general abilities. *Learn. Individ. Diff.* **39** (2015) 105–113.
- [14] D. Kalyanmoy and K. Miettinen, Nadir point estimation using evolutionary approaches: Better accuracy and computational speed through focused search. In: *Multiple Criteria Decision Making for Sustainable Energy and Transportation Systems*, edited by M. Ehrgott, B. Naujoks, T. Stewart and J. Wallenius. Springer Berlin Heidelberg, Berlin, Heidelberg (2010) 339–354.
- [15] M.L. Kern, L.E. Waters, A. Adler and M.A. White, A multidimensional approach to measuring well-being in students: Application of the perma framework. *J. Positive Psychol.* **10** (2015) 262–271.
- [16] G. Li, G. Hou, D. Yang, H. Jian and W. Wang, Relationship between anxiety, depression, sex, obesity, and internet addiction in chinese adolescents: A short-term longitudinal study. *Addict. Behav.* **90** (2019) 421–427.
- [17] M. Luque, F. Ruiz and K. Miettinen, Global formulation for interactive multiobjective optimization. *OR Spectr.* **33** (2011) 27–48.
- [18] M. Luque, O.D. Marcenaro-Gutiérrez and L.A. López-Agudo, On the potential balance among compulsory education outcomes through econometric and multiobjective programming analysis. *Eur. J. Oper. Res.* **241** (2015) 527–540.

- [19] L.C. Mann and M. Walshaw, Mathematics anxiety in secondary school female students: Issues, influences and implications. *N. Z. J. Educ. Stud.* **54** (2019) 101–120.
- [20] O.D. Marcenaro-Gutierrez, M. Luque and F. Ruiz, An application of multiobjective programming to the study of workers' satisfaction in the Spanish labour market. *Eur. J. Oper. Res.* **203** (2010) 430–443.
- [21] O.D. Marcenaro-Gutiérrez, M. Luque and L.A. López-Agudo, Balancing teachers' math satisfaction and other indicators of the education system's performance. *Soc. Indic. Res.* **129** (2016) 1319–1348.
- [22] N. Matthews, L. Kilgour, P. Christian, P. Mori and D.M. Hill, Understanding, evidencing, and promoting adolescent well-being: An emerging agenda for schools. *Youth Soc.* **47** (2015) 659–683.
- [23] K. Miettinen, *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston (1999).
- [24] OECD, PISA 2015 Results. Vol. III. OECD Publishing (2017) 1–528.
- [25] OECD, PISA 2015 Results in Focus. OECD Publishing (2018).
- [26] C.H. Oliveira and C.H. Antunes, Multiple objective linear programming models with interval coefficients – An illustrated overview. *Eur. J. Oper. Res.* **181** (2007) 1434–1463.
- [27] W.A. Spivey and H. Tamura, Goal programming in econometrics. *Nav. Res. Logist. Q.* **17** (1970) 183–192.
- [28] I.F. Strøm, S. Thoresen, T. Wentzel-Larsen and G. Dyb, Violence, bullying and academic achievement: A study of 15-year-old adolescents and their school environment. *Child Abuse Negl.* **37** (2013) 243–251.
- [29] M. Tamiz, D. Jones and C. Romero, Goal programming for decision making: An overview of the current state-of-the-art. *Eur. J. Oper. Res.* **111** (1998) 569–581.
- [30] H. Wallenius, J. Wallenius and P. Vartia, An approach to solving multiple criteria macroeconomic policy problems and an application. *Manage. Sci.* **24** (1978) 1021–1030.
- [31] P. Warr, Well-being and the workplace. In: *Well-Being: The Foundations of Hedonic Psychology*, edited by D. Kahneman, E. Diener and N. Schwarz. Russell Sage Foundation, New York (1999) 392–412.
- [32] A.P. Wierzbicki, The use of reference objectives in multiobjective optimization. In: *Multiple Criteria Decision Making, Theory and Applications*, edited by G. Fandel and T. Gal. Springer Berlin Heidelberg, Berlin, Heidelberg (1980) 468–486.

## Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

### Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting [subscribers@edpsciences.org](mailto:subscribers@edpsciences.org)

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/maths-s2o-programme>