University of Economics, Prague
Faculty of Informatics and Statistics
Department of Econometrics

Technical Report 05/16

Possibilistic linear regression with fuzzy data: Tolerance approach with prior information

Michal Černý and Milan Hladík

Preprint

October 2016

Possibilistic Linear Regression with Fuzzy Data: Tolerance Approach with Prior Information

Michal Černý[1] and Milan Hladík[2]

[1] *Department of Econometrics, University of Economics in Prague, Winston Churchill Square 4, 13067 Prague 3, Czech Republic, e-mail: cernym@vse.cz*

[2] *Department of Applied Mathematics, University of Mathematics and Physics, Charles University in Prague, Malostranské náměstí 25, 11000 Prague 1, Czech Republic, e-mail: milan.hladik@matfyz.cz.*

Abstract

We introduce the tolerance approach to the construction of fuzzy regression coefficients of a possibilistic linear regression model with fuzzy data (both input and output). The method is very general: the only assumption is that the fuzzy data are unimodal and their α-cuts are efficiently computable. We take into account possible prior restrictions of the parameters space: we assume that the restrictions are given by linear and quadratic constraints. The method for construction of the possibilistic regression coefficients is in a reduction of the fuzzy-valued model to an interval-valued model on a given α-cut, which is further reduced to a linear-time method computing with endpoints of the intervals. The speed of computation makes the method applicable for huge datasets.

Unlike various approaches based on mathematical programming formulations, the tolerance-based construction preserves central tendency of the resulting regression coefficients. In addition, we prove further properties: if inputs are crisp and outputs are fuzzy, then the construction preserves piecewise linearity and convex shape of fuzzy numbers. We derive an $O(n^2p)$-algorithm for enumeration of breakpoints of the membership function of the estimated coefficients. (Here, n is the number of observations and p is the number of regression parameters). Similar results are also derived for the fuzzy input-and-output model.

We illustrate the theory for the case of triangular and asymmetric Gaussian fuzzy inputs and outputs of an inflation-consumption model.

Keywords. Possibilistic regression; fuzzy regression; linear regression; constrained regression; tolerance quotient

1. Introduction

1.1. Traditional regression with crisp data. Notation.

General formulation. We consider the linear regression model

$$y_i = x_i^{\mathrm{T}}\beta + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where the vector of regression parameters $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is unknown and is to be estimated. The data $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ are called *outputs* and the data $x_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, $i = 1, \ldots, n$, are called *inputs*. The matrix $X = (x_1, \ldots, x_n)^{\mathrm{T}}$ is the matrix of *regressors*. The symbol ε_i stands for the random error.

Parameter space. The parameter space is denoted by $\mathfrak{B}$. It formalizes prior information about β: it is assumed that $\beta \in \mathfrak{B}$. The structure of $\mathfrak{B}$ usually arises from the physical, technical or economic meaning of regression parameters.

We consider either the *unconstrained case*, where $\mathfrak{B} = \mathbb{R}^p$, or the *constrained* case, where we assume constraints of the form

$$\mathfrak{B} = \{b \mid Ub \leq z,\ \|Cb - d\|_2 \leq g\}$$

with known U, z, C, d, g. (Inequalities are understood componentwise.) For example, when the regression parameters are a priori known to be nonnegative, then $Y = -I_{p\times p}$, $z = 0_{p\times 1}$ and C, b, g are empty.

A regression problem is then given by a triple $(y, X, \mathfrak{B})$.

1.2. The possibilistic regression concept with crisp data.

In traditional regression the task is to estimate the parameters $\beta \in \mathfrak{B}$ from the data (y, X). In possibilistic regression, the task is to compute a *feasible solution of the problem* (y, X), which is a set $B \subseteq \mathfrak{B}$ such that

$$(\forall i)(\exists b \in B)\, \underbrace{y_i = x_i^{\mathrm{T}} b}_{(\star)}. \tag{2}$$

In addition, the task is to find a feasible solution which is, in a sense, minimal, or “the best”; this will be formalized later in Sections 3.2 and 3.3.

The condition $(\star)$ tells us that the i-th data point (y_i, x_i) is *covered* by B.

Remark 1. *Possibilistic regression has been studied by several authors (see e.g. [21]) as a method complementary to traditional regression. The main goal of traditional regression is to capture the* average behavior *of the system modeled by (1); this is how estimators, such as least squares, are usually designed. Possibilistic regression belongs to the class of data envelopment methods, where the main goal is to capture also the best-case and worst-case behavior of the system under consideration. This viewpoint is useful e.g. in civil engineering — when building bridges, we usually prefer that they would not fall down even in the worst case, rather than on average.*

2. Possibilistic regression with fuzzy data: Problem formulation

2.1. Notation.

For a fuzzy number $\widetilde{a}$, let $\mu_{\widetilde{a}}$ denote its membership function. We restrict ourselves to the class of fuzzy numbers with unimodal membership functions. Then, $\mathrm{mod}_{\widetilde{a}}$ is the unique modus. We say that $\widetilde{a}$ is *piecewise linear* if $\mu_{\widetilde{a}}$ is a piecewise linear function. We say that $\widetilde{a}$ is *convex-shaped* if the function $\mu_{\widetilde{a}}$ is convex on the interval $(-\infty, \mathrm{mod}_{\widetilde{a}})$ as well as on the interval $(\mathrm{mod}_{\widetilde{a}}, \infty)$. We define $\widetilde{a} \subseteq \widetilde{b}$ if $\mu_{\widetilde{a}}(\xi) \leq \mu_{\widetilde{b}}(\xi)$ for all ξ.

Closed intervals are denoted in boldface. They can be represented either by endpoints or by the center and radius:

$$\boldsymbol{a} = [\underline{a}, \overline{a}] = [a^c \pm a^\Delta].$$

A similar notation $\boldsymbol{A}, \underline{A}, \overline{A}, A^c, A^\Delta$ is used for an interval matrix (or vector) $\boldsymbol{A}$. The space of $(m \times n)$ interval matrices is denoted by $\mathbb{IR}^{m \times n}$. Interval arithmetic is defined naturally [1, 19]:

$$\boldsymbol{a} \pm \boldsymbol{b} = [\underline{a} \pm \underline{b}, \overline{a} \pm \overline{b}], \quad \boldsymbol{a} \cdot \boldsymbol{b} = [\min(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}), \max(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b})].$$

The α-cut of a fuzzy number $\widetilde{a}$ is an interval denoted by $\boldsymbol{a}^\alpha = [\underline{a}^\alpha, \overline{a}^\alpha]$, where $\underline{a}^\alpha = \inf_\xi \mu_{\widetilde{a}}(\xi) \geq \alpha$ and $\overline{a}^\alpha = \sup_\xi \mu_{\widetilde{a}}(\xi) \geq \alpha$. We say that $\widetilde{a}$ is *bounded* if $\boldsymbol{a}^{\alpha=0}$ is bounded; otherwise $\widetilde{a}$ is *unbounded*. Arithmetic on fuzzy numbers is defined via α-cuts: for fuzzy numbers $\widetilde{a}, \widetilde{b}, \widetilde{c}$ and an operation $\circ \in \{+, \times\}$ we have $\widetilde{c} = \widetilde{a} \circ \widetilde{b}$ iff $\boldsymbol{c}^\alpha = \boldsymbol{a}^\alpha \circ \boldsymbol{b}^\alpha$ for every $\alpha \in (0, 1]$.

The absolute value $|A|$ of a real-valued matrix (vector) A is understood componentwise.

2.2. Feasibility.

We follow the traditional possibilistic approach to fuzzy regression. In the context of fuzzy data, this model was studied in [2, 3, 5, 7, 20, 21, 22, 25], among many others. We naturally generalize the feasibility concept to take into account the (possibly) restricted parameter space $\mathfrak{B}$.

First we define the notion of feasibility for the special case when X is crisp ("crisp-input-fuzzy-output model").

Definition 2. *A fuzzy vector $\widetilde{b}$ is a* feasible solution *of the possibilistic regression problem $(\widetilde{y}, X, \mathfrak{B})$ if*

$$\text{(i)} \;\; \widetilde{y} \subseteq X\widetilde{b} \quad \text{and} \quad \text{(ii)} \;\; \boldsymbol{b}^{\alpha=0} \subseteq \mathfrak{B}. \tag{3}$$

Remark 3. *We could have used a more general definition: a set is γ-*feasible *if $\widetilde{y} \subseteq X\widetilde{b}$ and $\boldsymbol{b}^\alpha \subseteq \mathfrak{B}$ for all $\alpha \geq 1 - \gamma$. Here, 1-feasibility is the same as feasibility in the sense of Definition 2. The γ-definition would not make any difference in the forthcoming theory; it is straightforward to generalize it from 1-feasibility to γ-feasibility and this is left to the reader.*

Condition (i) in (3) can be restated in terms of α-cuts:

$$(\forall\alpha)(\forall y \in \boldsymbol{y}^\alpha)(\forall i)(\exists b \in \boldsymbol{b}^\alpha)\ y_i = x_i^{\mathrm{T}} b,$$

showing that Definition 2 indeed generalizes (2) in the sense that "all data must be covered by $\widetilde{b}$, on every α-level".

When both $(\widetilde{y}, \widetilde{X})$ are fuzzy ("fuzzy-input-fuzzy-output model"), we make another natural step in generalization of the definition.

Definition 4 (fuzzy-input-fuzzy-output model)**.** *A fuzzy vector $\widetilde{b}$ is a* feasible solution *of the possibilistic regression problem $(\widetilde{y}, \widetilde{X}, \mathfrak{B})$ if*

$$\textit{(i)}\quad (\forall\alpha)(\forall y \in \boldsymbol{y}^\alpha)(\forall X \in \boldsymbol{X}^\alpha)(\forall i)(\exists b \in \boldsymbol{b}^\alpha)\ y_i = x_i^{\mathrm{T}} b \quad \textit{and} \quad \textit{(ii)}\ \ \boldsymbol{b}^{\alpha=0} \subseteq \mathfrak{B}.$$

Observe that Definition 2 is a special case of Definition 4 for X crisp. Definition 4 again follows the possibilistic paradigm and tells us that every possible data point $X \in \boldsymbol{X}^\alpha$ and $y \in \boldsymbol{y}^\alpha$ must be covered.

2.3. Organization of the paper.

We consider separately two questions: (i) testing whether a given solution is feasible for the model $(\widetilde{X}, \widetilde{y})$; and (ii) how to find a feasible solution $\widetilde{b}$ which is in a sense "the best one", or "the minimal one". Step (ii) will be done in terms of the tolerance approach in Section 3.3.

First we develop the tolerance approach for crisp-valued data (Section 3) and interval-valued data (Section 4). Then, its application to α-cuts of $(\widetilde{y}, \widetilde{X})$ yields the method for fuzzy data. We will prove several interesting properties of the constructed coefficients $\widetilde{b}$. Section 5 is devoted to the crisp-input-fuzzy-output model and Section 6 is devoted to the most general fuzzy-input-fuzzy-output model.

3. Crisp data

3.1. Feasibility.

We are to test whether a given $B \subseteq \mathbb{R}^p$ is feasible for a given crisp dataset (y, X). For the forthcoming theory it will be sufficient to restrict to the case when B is a p-dimensional interval (but the question is interesting for more general sets B, too).

Theorem 5. *Let $\mathfrak{B} = \{b \mid Ub \leq z,\ \|Cb - d\| \leq g\}$. Then an interval vector $\boldsymbol{b} \in \mathbb{IR}^p$ is feasible if and only if the following system is solvable:*

$$Xb^c - |X|b^\Delta \leq y \leq Xb^c + |X|b^\Delta, \tag{4a}$$

$$Ub^c + |U|b^\Delta \leq z, \tag{4b}$$

$$Cb^c + |C|b^\Delta - d \leq h, \tag{4c}$$

$$-Cb^c + |C|b^\Delta + d \leq h, \quad \|h\| \leq g. \tag{4d}$$

Proof. The constraint (4*a*) follows from the reformulation of $y \subseteq X\boldsymbol{b}$ as $y \in X\boldsymbol{b} = [Xb^c - |X|b^\Delta, Xb^c + |X|b^\Delta]$; see e.g. [1, 19].

The condition $Yb \le z\ \forall b \in \boldsymbol{b}$ can be reformulated as $Yb^c + |Y|b^\Delta \le z$ since again $Y\boldsymbol{b} = [Yb^c - |Y|b^\Delta, Yb^c + |Y|b^\Delta]$.

Eventually, the third condition $\|Cb - d\| \le g\ \forall b \in \boldsymbol{b}$ can be reformulated as $|Cb - d| \le h$, $\|h\| \le g\ \forall b \in \boldsymbol{b}$. Since $\max_{b\in\boldsymbol{b}} |Cb - d| = |Cb^c - d| + |C|b^\Delta$, the remaining constraints follows. □

Notice that (4) are linear and convex quadratic constraints, which are easily checked for solvability. Provided $\mathfrak{B}$ contains linear inequalities only, then all constraints in (4) are linear, too.

3.2. Finding a solution: General discussion.

If (4) is solvable, then it typically has infinitely many solutions. That is why a suitable solution must be chosen. In accordance with [14, 17, 20, 23, 25], we can utilize the linear programming problem

$$\min \sum_{i=1}^{p} b_i^\Delta \ \text{ subject to } \ (4), \tag{5}$$

or the convex quadratic programming problem

$$\min \sum_{i=1}^{p} (b_i^\Delta)^2 \ \text{ subject to } \ (4).$$

As it was several times mentioned [10, 16, 22, 23], these approaches suffer from several drawbacks. In particular, they often do not respect the *central tendency*. Central tendency is the following property: given a solution $\boldsymbol{b}$, its center b^c should provide a reasonable fit for data (y, X) with respect to the traditional goodness-of-fit measures, such as R-squared.

That is why we employ the successful tolerance approach developed in [10]: first, we fix the center b^c by traditional crisp-data fitting methods and then we extend the centers to intervals, assuring feasibility and minimality.

3.3. Tolerance approach.

Step 1. First we determine the center b^c, whose central tendency we want to respect, by traditional data fitting methods. For example, we can use the least squares $b^c = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$.

Step 2. Let a scaling vector $c \ge 0, c \ne 0$, called *tolerance vector*, be given by a user. We seek for $\boldsymbol{b}^\delta \in \mathbb{IR}^p$ in the form

$$\boldsymbol{b}^\delta = [b^c \pm \delta c],$$

The coefficient $\delta \ge 0$ is called *tolerance quotient*. We say that δ is *feasible* if $\boldsymbol{b}^\delta$ is feasible.

By introduction of the user-specified tolerance vector c we can control the widths of the resulting intervals $\boldsymbol{b}_i^{\delta}$. Typically, it is chosen as $c := (1,\dots,1)^{\mathrm{T}}$ for *absolute tolerances* and $c := |b^c|$ for *relative tolerances*. Nevertheless, it can be be set up in any other way depending on the significance of parameters and decision maker's preferences. For example, setting $c_{i_0} = 0$ forces $\boldsymbol{b}_{i_0}^{\delta}$ to be crisp; this is useful e.g. when the value $b_{i_0}^c$ is assumed to be known exactly.

Case I: $\mathfrak{B}$ is unconstrained. The minimal $\delta \geq 0$ such that all observations are covered is called *optimal*. It is denoted by δ^*, and is computed as follows.

Theorem 6 ([10])**.** *If there is $i \in \{1,\dots,n\}$ such that $|x_i^{\mathrm{T}}|c = 0$ and $y_i \neq x_i^{\mathrm{T}} b^c$ then there exists no feasible δ. Otherwise,*

$$\delta^* = \max_{i:|x_i^{\mathrm{T}}|c>0} \frac{|y_i - x_i^{\mathrm{T}} b^c|}{|x_i^{\mathrm{T}}|c}, \tag{6}$$

where $\max \emptyset = 0$ by definition. □

Case II: $\mathfrak{B}$ is constrained. For the constrained case, feasibility of $\boldsymbol{b}^{\delta^*}$, can be easily tested by Theorem 5.

Case III: $\mathfrak{B}$ is constrained only by linear constraints $Ub \leq z$. In this case, we do not need Theorem 5 in its full generality. We can easily compute the maximal $\delta^1 \geq 0$ such that $\boldsymbol{b}^{\delta^1} \subseteq \mathfrak{B}$ as

$$\delta^1 = \min_{i:|u_i^{\mathrm{T}}|c>0} \frac{y_i - u_i^{\mathrm{T}} b^c}{|u_i^{\mathrm{T}}|c}, \tag{7}$$

where u_i^{T} is the i-th row of U. It is easy to prove:

Theorem 7. *If $\delta^* \leq \delta^1$, then δ^* is feasible.* □

Remark 8. *If $\delta^* > \delta^1$, then we can try to extend $\boldsymbol{b}^{\delta^*}$ to a feasible interval vector $\boldsymbol{b} \supseteq \boldsymbol{b}^{\delta^1}$ such that central tendency and scaling vector is maximally respected. This is exactly the technique developed in [6, 9]. However, even this approach does not guarantee feasibility, so in that case is seems better to employ some optimization problem discussed in Section 3.2. Alternatively, we can try to find a minimal perturbation of c such that the tolerance approach will be solvable. This leads to the optimization problem*

$$\begin{aligned} \min_{\substack{b^{\Delta}\in\mathbb{R}^p\\ h\in\mathbb{R}^n}} \|c - b^{\Delta}\| \quad \text{subject to} \quad Xb - |X|b^{\Delta} \leq y \leq Xb + |X|b^{\Delta}&,\\ Ub + |U|b^{\Delta} \leq z&,\\ Cb + |C|b^{\Delta} - d \leq h&,\\ -Cb + |C|b^{\Delta} + d \leq h&, \quad \|h\| \leq g. \end{aligned}$$

Provided we use the Euclidean norm, this is a convex quadratic programming problem. Provided we employ ℓ_1 or ℓ_∞-norm and there are no quadratic constraints in $\mathfrak{B}$, then it is a linear program.

4. Interval-valued data

Interval regression was intensively investigated in the recent decades. The possibilistic concept for interval data was originally developed by Tanaka et al. [24] (see also [20, 23]) and it has been applied in a number of practical problems.

There are different methods used for possibilistic regression. The basic one employs a linear programming formulation [14, 17, 20, 25]. It is quite simple and requires small computational effort. However, this method suffers from several undesirable properties. Often some of the estimated regression parameters tend to be crisp while it simultaneously produces a few of unexpectedly wide interval parameters [12, 13, 16, 23]. The second undesirable property is non-centrality discussed in Section 3.2.

To overcome these drawbacks, various alternative methods have been proposed. A quadratic programming model was proposed in [23]. Recently, a lot of effort was done in interval regression analysis using support vector machines [4, 8, 11, 12, 13, 15]. In [10], a tolerance based approach was presented: It not only overcomes the above mentioned drawbacks, but also is efficiently computable. That is why our method is based on this approach.

First we consider the *crisp-input-interval-output* case $(\boldsymbol{y}, X)$. Then we turn the *interval-input-interval-output* case $(\boldsymbol{y}, \boldsymbol{X})$.

4.1. Crisp-input-interval-output model

Feasibility. Let a dataset $(\boldsymbol{y}, X)$ be given with interval $\boldsymbol{y} = [\underline{y}, \overline{y}]$. Now, an interval vector $\boldsymbol{b} \in \mathbb{IR}^p$ is *feasible* for the problem $(\boldsymbol{y}, X)$ if $(\forall i)(\forall y_i \in \boldsymbol{y}_i)(\exists b \in \boldsymbol{b})\ y_i = x_i^{\mathrm{T}} b$, or $\boldsymbol{y} \subseteq X\boldsymbol{b}$ for short. As in Theorem 5, feasibility can be easily verified.

Theorem 9. *Let $\mathfrak{B} = \{b \mid Ub \le z,\ \|Cb - d\| \le g\}$. The interval vector $\boldsymbol{b}$ is feasible iff*

$$Xb^c - |X|b^\Delta \le \underline{y}, \tag{8a}$$

$$Xb^c + |X|b^\Delta \ge \overline{y}, \tag{8b}$$

$$Ub^c + |U|b^\Delta \le z, \tag{8c}$$

$$Cb^c + |C|b^\Delta - d \le h, \tag{8d}$$

$$-Cb^c + |C|b^\Delta + d \le h, \quad \|h\| \le g. \tag{8e}$$

Proof. Analogous to Theorem 5. □

Finding a solution: General discussion. One of the first approaches [14, 17, 20, 25] was to solve the LP problem

$$\min_{b^c, b^\Delta} \sum_{i=1}^{n} |x_i^{\mathrm{T}}|b^\Delta \quad \text{s.t.} \quad \underline{y}_i \ge x_j^{\mathrm{T}} b^c - |x_j^{\mathrm{T}}|b^\Delta, \quad i = 1, \ldots, n,$$

$$\overline{y}_j \le x_j^{\mathrm{T}} b^c + |x_j^{\mathrm{T}}|b^\Delta, \quad i = 1, \ldots, n,$$

and it is not surprising that the non-centrality problem is here analogous to the discussion of Section 3.2. Again, we resolve it by the tolerance approach.

Tolerance approach. Let a crisp estimate b^c be given; say for example that it results from centralized least squares $b^c = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y^c$. Let the vector of tolerances $c \geq 0$, $c \neq 0$ be given. We are to find the minimum tolerance quotient δ^* such that $\boldsymbol{b}^{\delta^*} = [b^c \pm \delta^* c]$ is feasible for $(X, \boldsymbol{y})$. This question can be reduced to the crisp-data case. (We do not discuss feasibility of $\boldsymbol{b}^{\delta^*}$ in $\mathfrak{B}$ any more since feasibility can be tested by Theorem 9.)

Theorem 10 ([10]). *For $i = 1, \ldots, n$ let*

$$y_i := \begin{cases} \underline{y}_i & \text{if } |\underline{y}_i - x^{\mathrm{T}} b^c| \geq |\overline{y} - x_i^{\mathrm{T}} b^c|, \\ \overline{y}_i & \text{otherwise.} \end{cases}$$

Compute δ^* from (6) with data (y, X). Then, δ^* is the optimal tolerance quotient for $(\boldsymbol{y}, X)$. □

4.2. Interval-input-interval-output model

Feasibility. In the possibilistic approach to regression, feasibility corresponds to existence of an interval vector $\boldsymbol{b} \in \mathbb{IR}^p$ such that $(\forall i)(\forall X \in \boldsymbol{X})(\forall y \in \boldsymbol{y})(\exists b \in \boldsymbol{b})\ y_i = x_i^{\mathrm{T}} b$. Said informally, $\boldsymbol{b}$ must cover every data point $X \in \boldsymbol{X}$ and $y \in \boldsymbol{y}$.

Theorem 11. *Let $\mathfrak{B} = \{b \mid Ub \leq z,\ \|Cb - d\| \leq g\}$. Then an interval vector $\boldsymbol{b}$ is feasible if and only if*

$$\overline{y}_i \leq \sum_j d_{ij}, \tag{9a}$$
$$d_{ij} \leq \overline{X}_{ij} b_j^c - |\overline{X}_{ij}| b_j^{\Delta}, \tag{9b}$$
$$d_{ij} \leq \underline{X}_{ij} b_j^c - |\underline{X}_{ij}| b_j^{\Delta}, \tag{9c}$$
$$d_{ij} \leq 0 \quad \text{if } 0 \in \boldsymbol{X}_{ij}, \tag{9d}$$
$$\underline{y}_i \geq \sum_j g_{ij}, \tag{9e}$$
$$g_{ij} \geq \overline{X}_{ij} b_j^c + |\overline{X}_{ij}| b_j^{\Delta}, \tag{9f}$$
$$g_{ij} \geq \underline{X}_{ij} b_j^c + |\underline{X}_{ij}| b_j^{\Delta}, \tag{9g}$$
$$g_{ij} \geq 0 \quad \text{if } 0 \in \boldsymbol{X}_{ij}, \tag{9h}$$
$$Ub^c + |U| b^{\Delta} \leq z, \tag{9i}$$
$$Cb^c + |C| b^{\Delta} - d \leq h, \tag{9j}$$
$$-Cb^c + |C| b^{\Delta} + d \leq h, \quad \|h\| \leq g, \tag{9k}$$

is feasible.

Proof. The interval vector $\boldsymbol{b}$ is feasible if and only if $\boldsymbol{b} \subseteq \mathfrak{B}$ and

$$\boldsymbol{y} \subseteq X\boldsymbol{b} \quad \forall X \in \boldsymbol{X}.$$

The latter can be expressed as

$$\overline{y} \le \underline{X\boldsymbol{b}},\ \underline{y} \ge \overline{X\boldsymbol{b}}, \quad \forall X \in \boldsymbol{X}.$$

Since $\underline{X\boldsymbol{b}} = Xb^c - |X|b^\Delta$, we can write the first inequality as $\overline{y}_i \le Xb^c - |X|b^\Delta\ \forall X \in \boldsymbol{X}$. The function $Xb^c - |X|b^\Delta$ is piecewise linear, so its minimum is attained at $X_{ij} \in \{\overline{X}_{ij}, \underline{X}_{ij}, 0\}$. Introducing an auxiliary vector of variables d_i, we can equivalently write the inequality as

$$\overline{y}_i \le e^{\mathrm{T}} d_i,\ d_{ij} \le \overline{X}_{ij} b^c_j - |\overline{X}_{ij}| b^\Delta_j,\ d_{ij} \le \underline{X}_{ij} b^c_j - |\underline{X}_{ij}| b^\Delta_j,\ d_{ij} \le 0 \text{ if } 0 \in \boldsymbol{X}_{ij},$$

where $e = (1, \dots, 1)^T$. Similarly we proceed for the other inequalities. Eventually, the condition $\boldsymbol{b} \subseteq \mathfrak{B}$ is handled as in the proof of Theorem 5. ☐

The system (9) is linear, so its feasibility can by checked by means of linear programming. Notice, however, that the system has a large number of variables (d_{ij}, g_{ij}, h_k and possibly b^c, b^Δ, similarly to (5)). That is why a computationally cheaper method might be desirable. Below, we propose simpler approach, but we pay for the lower computational cost by strength since it works as a sufficient condition only.

Theorem 12. *Let $\mathfrak{B} = \{b \mid Ub \le z,\ \|Cb - d\| \le g\}$. Then an interval vector $\boldsymbol{b}$ is feasible if*

$$X^c b^c - X^\Delta |b^c| - (|X^c| + X^\Delta) b^\Delta \ge \overline{y}, \tag{10a}$$

$$X^c b^c + X^\Delta |b^c| + (|X^c| + X^\Delta) b^\Delta \le \underline{y}, \tag{10b}$$

$$Ub^c + |U| b^\Delta \le z, \tag{10c}$$

$$Cb^c + |C| b^\Delta - d \le h, \tag{10d}$$

$$-Cb^c + |C| b^\Delta + d \le h, \quad \|h\| \le g, \tag{10e}$$

is feasible.

Proof. We already know that the interval vector $\boldsymbol{b}$ is feasible if and only

$$\overline{y} \le Xb^c - |X|b^\Delta,\ \underline{y} \ge Xb^c + |X|b^\Delta, \quad \forall X \in \boldsymbol{X}.$$

Utilizing the estimates

$$Xb^c \ge X^c b^c - X^\Delta |b^c|, \quad |X| b^\Delta \le (|X^c| + X^\Delta) b^\Delta,$$

we replace $\overline{y} \le Xb^c - |X|b^\Delta$ by the stronger condition

$$\overline{y} \le X^c b^c - X^\Delta |b^c| - (|X^c| + X^\Delta) b^\Delta.$$

Analogously we process the remaining constraints. ☐

The system (10) is obviously much smaller than (9). On the other hand, if b^c, b^Δ serve as variables, then the system (10) is nonlinear due to the absolute value. Good news is that we can equivalently reformulate it avoiding the absolute value:

$$\begin{aligned}
X^c b^c - X^\Delta d - (|X^c| + X^\Delta) b^\Delta &\geq \overline{y}, \\
X^c b^c + X^\Delta d + (|X^c| + X^\Delta) b^\Delta &\leq \underline{y}, \\
b^c &\leq d, \\
-b^c &\leq d, \\
U b^c + |U| b^\Delta &\leq z, \\
C b^c + |C| b^\Delta - d &\leq h, \\
-C b^c + |C| b^\Delta + d &\leq h, \quad \|h\| \leq g.
\end{aligned}$$

Now we have a linear system consisting of only $3p$ variables (b^c, b^Δ, d).

Tolerance approach. In [10], the authors designed the following reduction of the "interval-input, interval-output" case to the crisp case. Notice that it can slightly overestimate the optimal interval vector $\boldsymbol{b}$, but it is efficient. It can be proved the overestimation does not occur in some natural situations, e.g., when

$$0 \notin \boldsymbol{b}_i, \quad i = 1, \ldots, p. \tag{11}$$

Let the tolerance vector $c \geq 0, c \neq 0$ and the central estimator b^c be given (for example, b^c can be obtained from the central least squares $b^c = [(X^c)^{\mathrm{T}} X^c]^{-1} (X^c)^{\mathrm{T}} y^c$). The method constructs two auxiliary crisp-input-crisp-output models

$$(y^1, X^1) \quad \text{and} \quad (y^2, X^2)$$

and computes their optimal tolerance quotients δ^1, δ^2, respectively, where

$$\begin{aligned}
y_i^1 &:= \underline{y}_i, \quad x_{ij}^1 = \begin{cases} \overline{x}_{ij} & \text{if } b_i^c \geq 0, \\ \underline{x}_{ij} & \text{if } b_i^c < 0, \end{cases} \\
y_i^2 &:= \overline{y}_i, \quad x_{ij}^2 = \begin{cases} \underline{x}_{ij} & \text{if } b_i^c \geq 0, \\ \overline{x}_{ij} & \text{if } b_i^c < 0 \end{cases}
\end{aligned}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. The resulting tolerance quotient $\delta^* = \max\{\delta^1, \delta^2\}$ is minimal only under additional assumptions, such as the sign invariancy assumption (11).

5. Fuzzy-valued data: Crisp-input-fuzzy-output model

First we study the case of crisp-input data $(\widetilde{y}, X)$. We apply the tolerance approach developed for interval-output data on the level of α-cuts of $\widetilde{y}$.

Again we assume that the central crisp estimate b^c of regression coefficients and the tolerance vector $c \geq 0$, $c \neq 0$ are given. We are to construct a fuzzy vector $\widetilde{b}$, which is a feasible solution of $(\widetilde{y}, X)$ and is of the form

$$\boldsymbol{b}^\alpha = [b^c \pm \delta^*(\alpha) c]$$

for every α, where $\delta^*(\alpha)$ is the optimal tolerance quotient for the interval-output problem $(\boldsymbol{y}^\alpha, X)$.

Discussion: how to choose b^c. To get the initial crisp-value estimate b^c, any usual defuzzification method for $\widetilde{y}$ can be used, e.g. taking the centroids y^*. Then, crisp-data estimators for (y^*, X) yield the initial value b^c.

5.1. Properties of the fuzzy regression coefficients.

To prove correctness—that $\widetilde{b}$ is a well-defined unimodal fuzzy number—it suffices to observe that $\delta^*(\alpha)$ is a non-increasing function of α.

The following property shows that whenever we work in the class of piecewise linear fuzzy numbers, then our method produces also fuzzy regression coefficients in this class.

Theorem 13. *If the fuzzy numbers in $\widetilde{y}$ are piecewise linear, then so are the fuzzy numbers in $\widetilde{b}$.*

Proof. We must prove that $\delta^*(\alpha)$ is a piecewise linear function of α. The absolute value in (6) can be w.l.o.g. omitted, depending on the sign of $y_i - x_i^{\mathrm{T}} b^c$. Therefore, if the endpoints $\underline{y}_i^\alpha, \overline{y}_i^\alpha$ of α-cuts of $\widetilde{y}$ are piecewise linear functions of α, then δ^* is a maximum of linear functions, so it is piecewise linear. $\square$

A similar property holds for the class of convex-shaped fuzzy numbers.

Theorem 14. *If the fuzzy numbers in $\widetilde{y}$ are convex-shaped, then so are the fuzzy numbers in $\widetilde{b}$.*

Proof. Similarly as in the proof of Theorem 13, the absolute value in (6) can be omitted, depending on the sign of $y_i - x_i^{\mathrm{T}} b^c$. Therefore, if the endpoints $\underline{y}_i^\alpha, \overline{y}_i^\alpha$ are piecewise linear functions of α are convex functions of α, then δ^* is maximum of convex functions, so it is convex as well. $\square$

As a consequence of the above results, when the fuzzy numbers in $\widetilde{y}$ are triangular, then the fuzzy numbers in $\widetilde{b}$ are piecewise linear and convex-shaped. The following results shows that it may happen that even the elements in $\widetilde{b}$ are triangular provided some strong but quite natural assumptions hold.

Recall that the symmetric triangular fuzzy number $\widetilde{y} = (y - y^\Delta, y, y + y^\Delta)$ is defined by its α-cut as $\boldsymbol{y}^\alpha = [y \pm (1-\alpha) y^\Delta]$.

Theorem 15. *If $\widetilde{y}_i$ is triangular $\widetilde{y}_i = (y_i - y_i^\Delta, y_i, y_i + y_i^\Delta)$ with radius $y_i^\Delta = |x_i^{\mathrm{T}}| c$, then the elements of $\widetilde{b}$ are triangular, too.*

Proof. According to the formula (6), the function $\delta^*(\alpha)$ is maximum of linear functions. By the assumptions, all these linear functions have the same slope, so they cannot intersect (or coincide). Therefore, the maximum is attained for a certain i constantly for all $\alpha \in [0, 1]$. $\square$

5.2. Piecewise linear fuzzy numbers: Explicit construction of the fuzzy regression parameters.

Suppose that the fuzzy numbers in $\widetilde{y}$ are piecewise linear. As we observed, the fuzzy numbers in $\widetilde{b}$ are piecewise linear, too. We show a construction how to compute the break points of the piecewise linear function $\mu_{\widetilde{b}_i}$. For the sake of simplicity, we present the construction for the case $\widetilde{y}_i$ is a triangular fuzzy number $\widetilde{y}_i = (y_i - y_i^\Delta, y_i, y_i + y_i^\Delta)$; a generalization is straightforward.

Suppose that $y \geq Xb^c$; if $y_i < x_i^{\mathrm{T}} b^c$ for some i, then multiply the ith row by -1 (this is just for the simplification of exposition, no change of data is necessary). Then $\delta^*(\alpha)$ can be expressed as

$$\delta^*(\alpha) = \max_{i:|x_i^{\mathrm{T}}|c>0} \frac{\alpha y_i^\Delta + y_i - x_i^{\mathrm{T}} b^c}{|x_i^{\mathrm{T}}|c}.$$

Let the maximum be attained at i^* for $\alpha = 1$; if there are more possibilities, take that one with the largest slope $y_i^\Delta/(|x_i^{\mathrm{T}}|c)$. If i^* is not the maximizer for all $\alpha \in [0, 1]$, then there must exist the equilibrium

$$\frac{\alpha y_i^\Delta + y_i - x_i^{\mathrm{T}} b^c}{|x_i^{\mathrm{T}}|c} = \frac{\alpha y_{i^*}^\Delta + y_{i^*} - x_{i^*}^{\mathrm{T}} b^c}{|x_{i^*}^{\mathrm{T}}|c}$$

for some $i \neq i^*$. By an easy algebraic treatment we have that the equilibrium is attained for

$$\alpha = \frac{(y_i - x_i^{\mathrm{T}} b^c)|x_{i^*}^{\mathrm{T}}|c - (y_{i^*} - x_{i^*}^{\mathrm{T}} b^c)|x_i^{\mathrm{T}}|c}{y_{i^*}^\Delta |x_i^{\mathrm{T}}|c - y_i^\Delta |x_{i^*}^{\mathrm{T}}|c}. \tag{12}$$

As long as $\alpha \notin (0, 1)$, we do not need to consider i any more. So for all admissible $i \neq i^*$ we take that one which corresponds to the maximal α. This gives us the break point of the fuzzy numbers in $\widetilde{b}$, and analogously we proceeds further until we come up to $\alpha = 0$.

The computational cost is $\mathcal{O}(n^2 p)$ since in the worst case we have to compute α from (12) for each pair of indices $j, j' \in \{1, \ldots, n\}$. Denoting ℓ the number of resulting break-points, we can express the computational cost also as $\mathcal{O}(n\ell p)$. Since the number of break-points is usually very mild, this bound gives more accurate estimation.

5.3. Discussion: Other approaches to the construction of $\widetilde{b}$.

In general, it is hard to say anything about the shape of the fuzzy regression parameters computed by more complicated optimization problems as other authors often do (recall [14, 17, 20, 23, 25]). Nevertheless, it can be shown that piecewise linearity of $\widetilde{b}$ (assuming piecewise linearity of $\widetilde{y}$) is still preserved when α-cuts of $\widetilde{b}$ are expressible as linear programs (5). (But note that these α-cuts need not be nested; it means that generally it is not guaranteed that (5) produces a well-defined fuzzy numbers in $\widetilde{b}$.)

x_i	1	2	3	4	5	6	7	8
y_i^1 (pessimistic)	12	10	15	15	25	30	45	70
y_i^2 (modal)	25	17	28	28	45	60	55	95
y_i^3 (optimistic)	29	28	38	60	60	65	90	100

Table 1: Data for Section 5.4.

Theorem 16. *If the fuzzy numbers in $\widetilde{y}$ are piecewise linear, then so are the fuzzy numbers in $\widetilde{b}$ provided α-cut interval regression problems are computed by* (5) *and the result gives well-defined fuzzy numbers in $\widetilde{b}$.*

Proof. Denote the system (4) as $Az \leq a$, where z is a vector of variables containing b^c and b^Δ, A is the constraint matrix, and a is the right-hand side vector containing y, among others. It is known that an optimal solution is attained at a vertex taking the form of $z = A_B^{-1} a_B$, where B is an optimal basis and the expressions A_B, a_B stand for the restriction to basic rows. Now, when $\underline{y}^\alpha, \overline{y}^\alpha$ are piecewise linear functions of α, then so is z and therefore an optimal solution also moves piecewise linearly. (The breakpoint can be also occur when a basis is changed.) Since the midpoint and the radius of $\boldsymbol{b}^\alpha$ changes piecewise linearly, the fuzzy vector $\widetilde{b}$ has piecewise linear shape. □

Since the breakpoints can occur when an optimal basis is changed, their number can potentially be very high (or, at least the proof does not indicate a nontrivial bound).

5.4. Example.

In this section we illustrate the tolerance approach by adaptation of an example from [23], where the quadratic trend

$$y_i = b_1 + b_2 x_i + b_3 x_i^2, \quad i = 1, \dots, n := 8$$

is modeled. Here, $x_i = 1, \dots, 8$ are crisp inputs and the values of the output variable y are expert-given: each value is given as a triplet y_i^1, y_i^2, y_i^3, corresponding to a pessimistic, modal and optimistic estimate of outcome. Particular values of this example are summarized in Table I.

Model I: We model $\widetilde{y}_i$ as the triangular fuzzy number (y_i^1, y_i^2, y_i^3).

Model II: We model $\widetilde{y}_i$ as the asymmetric Gaussian fuzzy number with membership

$$\mu_{\widetilde{y}_i}(\xi) = \begin{cases} \exp\left(-\frac{(\xi - y_i^2)^2}{\frac{1}{2}(y_i^2 - y_i^1)^2}\right) & \text{if } \xi \leq y_i^2, \\ \exp\left(-\frac{(\xi - y_i^2)^2}{\frac{1}{2}(y_i^3 - y_i^2)}\right) & \text{if } \xi > y_i^2. \end{cases}$$

The data are depicted in Figure 1, together with the reference parabola $y = b_1^c + b_2^c x + b_3^c x^2$, where

$$b^c = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} y^2 = (26.91, -5.72, 1.68)^{\mathrm{T}}$$

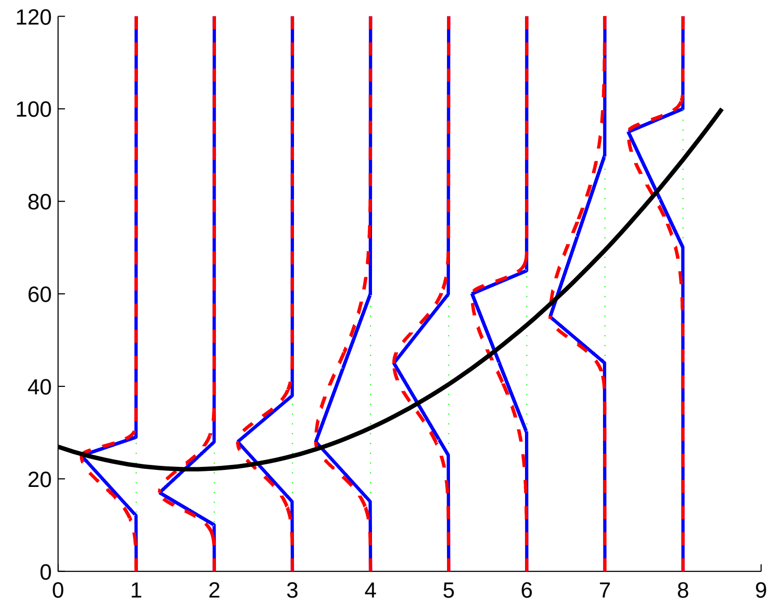

Figure 1: Example 5.4. Output variable modeled as a triangular fuzzy numbers (solid lines) and asymmetric Gaussian fuzzy numbers (dashed lines), and the central parabolic trend based on modal values y^2.

was calculated by the least squares from the modal estimates y^2.

Relative tolerances. The choice $c = |b^c|$ is called *relative tolerance*; here, the widths should be proportional to the absolute values of b^c.

Absolute tolerances. The choice $c = (1, 1, 1)^{\mathrm{T}}$ is called *absolute tolerance*; here, the widths have equal weights.

The resulting membership functions of $\widetilde{b}$, for both absolute and relative tolerances, are depicted in Figure 2 (solid lines: triangular $\widetilde{y}_i$, dashed lines: Gaussian $\widetilde{y}_i$). Recall that they have the following property: given α, *the α-cut of $\widetilde{b}$ covers the α-cut of $\widetilde{y}_i$ for all i* and is the minimal one with this property with respect to the prescribed tolerance vector c. Moreover, the central tendency around b^c — the natural least-squares estimator — is guaranteed.

Moreover, we can see that the membership functions are piecewise linear and convex-shaped if $\widetilde{y}_i$ are triangular (see Theorems 13, 14).

Restriction of the parameter space. The Gaussian fuzzy numbers are unbounded. But the triangular fuzzy numbers are bounded; we can see that in the triangular case, $\widetilde{b}$ are bounded by

$$\boldsymbol{b}^{\alpha=0}_{\text{relative}} = [16.4, 37.1] \times [-7.88, -3.55] \times [1.04, 2.32]$$

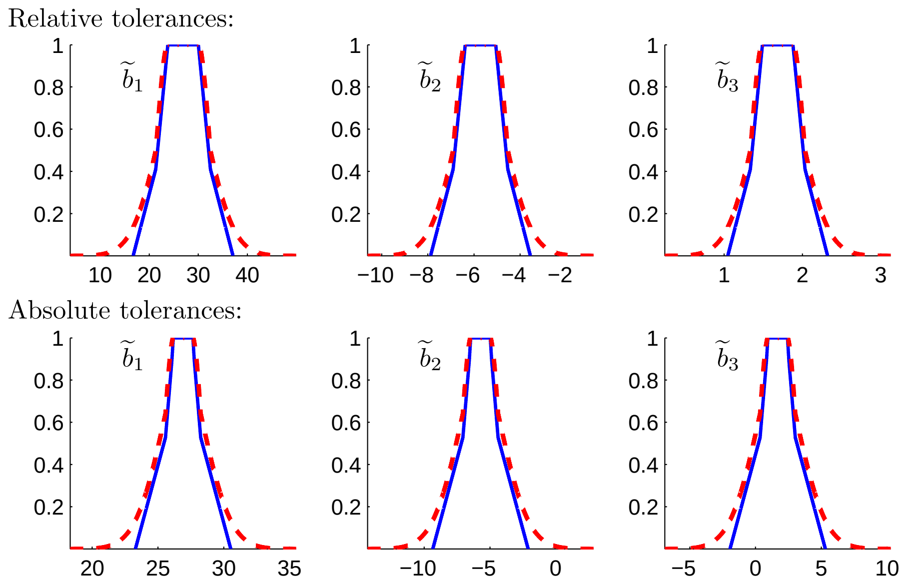

Figure 2: Example 5.4. The fuzzy regression coefficients $\widetilde{b}$ for relative tolerances (upper triplet) and absolute tolerances (bottom triplet). The results for triangular $\widetilde{y}_i$ are depicted by solid lines and the results for Gaussian $\widetilde{y}_i$ are depicted by dashed lines.

for relative tolerances and by

$$\boldsymbol{b}^{\alpha=0}_{\text{absolute}} = [23.29, 30.54] \times [-9.34, -2.10] \times [-1.94, 5.31]$$

for absolute tolerances. Recall that if the parameter space $\mathfrak{B}$ is restricted, then the model is said to be feasible if $\boldsymbol{b}^{\alpha=0} \subseteq \mathfrak{B}$. The most natural restriction of our model is $\mathfrak{B} = \{b \in \mathbb{R}^3 \mid b_3 \geq 0\}$. We can immediately see that the model with relative tolerances is feasible, while the model with absolute tolerances is not. This shows that the former model is more suitable than the latter.

6. Fuzzy-valued data: Fuzzy-input-fuzzy-output model

Now we consider the most general case $(\widetilde{y}, \widetilde{X})$ where all data can be fuzzy. Recall that the notion of feasibility of $\widetilde{b}$ w.r.t. $(\widetilde{y}, \widetilde{X}, \mathfrak{B})$ was formalized in Definition 4.

6.1. Tolerance approach.

We construct the vector $\widetilde{b}$ by reduction to the interval-input-interval-output case via α-cuts.

Assume again that we are given an initial crisp estimate b^c of regression coefficients (say, obtained by crisp least squares from defuzzified $(\widetilde{y}, \widetilde{X})$), whose

i	1	2	3	4	5	6	7	8	9	10
y_i (consumption, crisp)	39	49	30	45	43	54	39	66	60	53
M_i (income, crisp)	44	48	50	50	51	60	65	66	66	70
π_i^1 (exp. infl.: optimistic)	3.7	3.8	2.3	6.0	3.9	4.6	-1.6	5.2	6.2	2.2
π_i^2 (exp. infl.: modal)	4.0	6.0	3.0	7.0	5.0	6.0	1.0	7.0	8.0	4.0
π_i^3 (exp. infl.: pessimistic)	4.3	7.5	5.1	11.5	5.3	6.3	1.3	7.3	9.6	7.3
i	11	12	13	14	15	16	17	18	19	20
y_i (consumption, crisp)	55	45	63	41	59	69	54	60	51	73
M_i (income, crisp)	72	80	83	85	85	85	90	99	101	120
π_i^1 (exp. infl.: optimistic)	8.7	0.0	1.6	0.1	1.1	0.6	2.5	3.2	0.9	3.1
π_i^2 (exp. infl.: modal)	9.0	2.0	3.0	1.0	5.0	6.0	3.0	4.0	2.0	6.0
π_i^3 (exp. infl.: pessimistic)	11.6	3.2	3.3	2.7	5.3	6.3	4.3	4.3	2.3	7.2

Table 2: Data for regression model (13).

central tendency should be preserved, and a tolerance vector $c \geq 0$, $c \neq 0$. We are to construct the fuzzy vector $\widetilde{b}$, which is a feasible solution of $(\widetilde{y}, \widetilde{X})$, and is of the form

$$\boldsymbol{b}^\alpha = [b^c \pm \delta^*(\alpha)c]$$

for every α, where where $\delta^*(\alpha)$ is the optimal tolerance quotient for the interval regression problem $(\boldsymbol{y}^\alpha, \boldsymbol{X}^\alpha)$.

What can be said about the shape of the resulting fuzzy vector $\widetilde{b}$ entries?

Theorem 17. *If the fuzzy numbers in $\widetilde{y}, \widetilde{X}$ are piecewise linear, then the fuzzy numbers in $\widetilde{b}$ are piecewise hyperbolic.*

Proof. For each $\alpha \in [0, 1]$, the α-cut reduces the problem to interval-input-interval-output regression. The tolerance quotient is then computed by a formula of type (6) since even the interval regression is solved by a reduction to a double crisp regression model. Clearly, $\underline{y}_i^\alpha, \overline{y}_i^\alpha, \underline{x}_{ij}^\alpha, \overline{x}_{ij}^\alpha$ depend piecewise linearly on α, so the function (6) is a maximum of fractions of piecewise linear functions. Geometrically it is of a piecewise hyperbolic shape. □

6.2. Application.

It has been repeatedly demonstrated that consumption expenditures of households are better explained by *subjectively perceived* inflation or inflation expectations, rather than the true reported inflation by official statistics or inflation predictions published by authorities, banks or experts. (More on measurement of subjective expectations can be found e.g. in [18].) Subjectively perceived inflation often differs significantly from the official inflation rate because different households have a different structure of consumption. For example, regular suburban commuters are often more sensitive to inflation driven by prices of fuels (since fuels have a significant weight in their individual market basket) than urban residents. Another example: a household may be sensitive to inflation driven by currency deprecitation if imports have a significant weight in its individual market basket, while other households may be insensitive to changes in

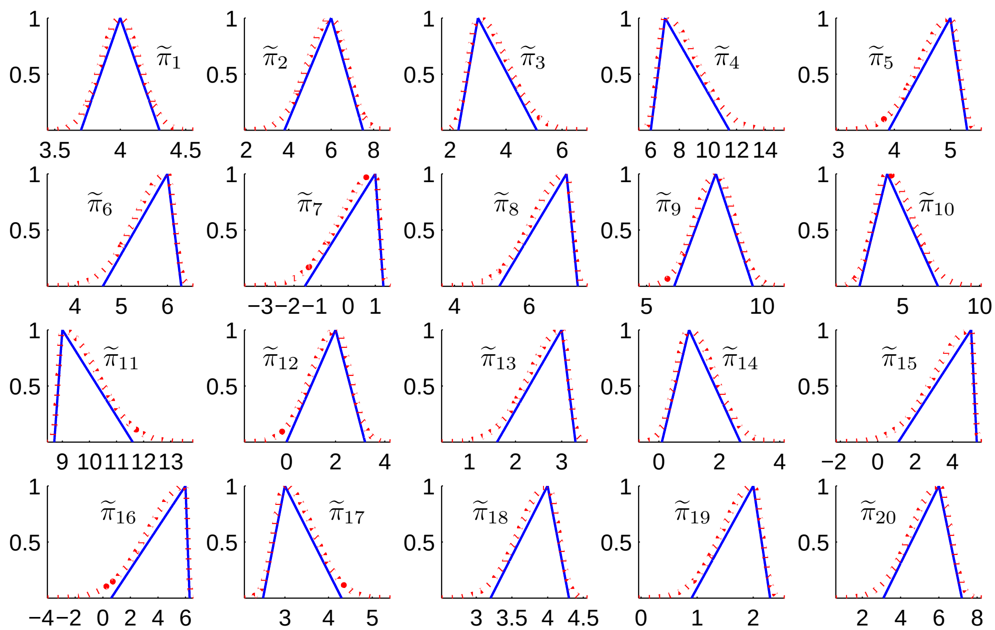

Figure 3: Membership functions of fuzzy input data (Section 6.2): subjectively perceived inflation expectations $\widetilde{\pi}_i$ modeled as triangular fuzzy numbers (solid lines) and asymmetric Gaussian fuzzy numbers (dotted lines).

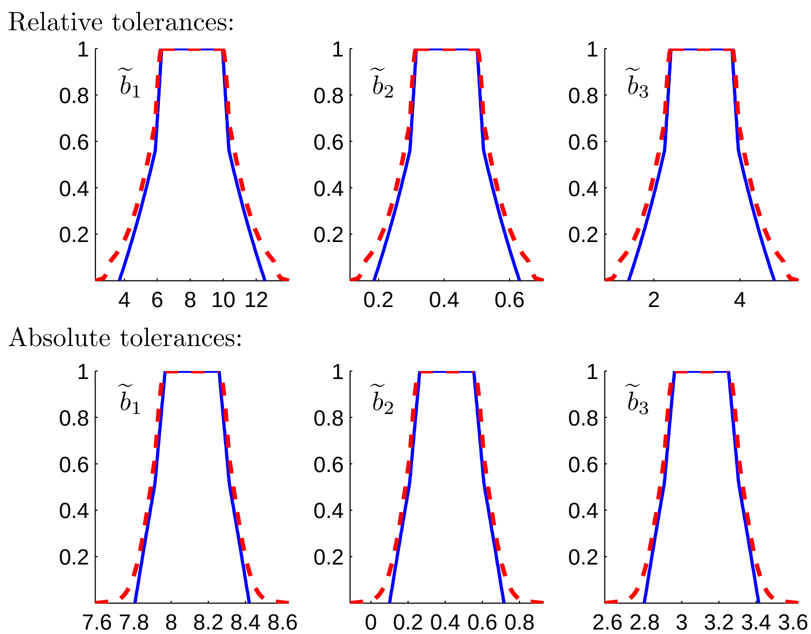

Figure 4: Fuzzy regression coefficients $\widetilde{b}$ for model (13) with triangular inputs $\widetilde{\pi}_i$ (solid line) and Gaussian inputs $\widetilde{\pi}_i$ (dashed line).

exchange rates. Generally, the structure of consumption differs among regions, age cohorts and also reflect individual optimism of pessimism.

We consider the cross-sectional model

$$y_i = \beta_1 + \beta_2 M_i + \beta_3 \pi_i + \varepsilon_i, \tag{13}$$

where we have a sample of $n = 20$ households. For ith household, y_i stands for consumption expenditures, M_i stands for disposable income and π_i stands for subjective perception of future inflation. The data y_i and M_i are crisp. The subjective inflation π_i has been measured as a triplet of values $(\pi_i^1, \pi_i^2, \pi_i^3)$: each household has been asked to give an optimistic (π_i^1), modal (π_i^2) and pessimistic (π_i^3) subjective prediction of inflation for the next period.

Model I. We model π_i as the triangular fuzzy number $\widetilde{\pi}_i = (\pi_i^1, \pi_i^2, \pi_i^3)$.

Model II. We model π_i as the asymmetric Gaussian fuzzy number $\widetilde{\pi}_i$ with

$$\mu_{\widetilde{\pi}_i}(\xi) = \begin{cases} \exp\left(-\frac{(\xi-\pi_i^2)^2}{\frac{1}{2}(\pi_i^2-\pi_i^1)^2}\right) & \text{if } \xi \leq \pi_i^2, \\ \exp\left(-\frac{(\xi-\pi_i^2)^2}{\frac{1}{2}(\pi_i^3-\pi_i^2)^2}\right) & \text{if } \xi > \pi_i^2. \end{cases} \tag{14}$$

Data are summarized in Table II and the membership functions of $\widetilde{\pi}_1, \ldots, \widetilde{\pi}_n$ are depicted in Figure 3 for both Models I and II.

The initial crisp estimate

$$b^c = (8.11, 0.41, 3.1)^{\mathrm{T}}$$

was obtained by least squares from modal values π_i^2. It shows that on average, each point of subjectively perceived inflation contributes to consumption expenditures by \$3.1. Now we construct the fuzzy regression coefficients $\widetilde{b}$ with two choices: relative tolerances $c = |b^c|$ $(= b^c)$ and absolute tolerances $c = (1, 1, 1)^{\mathrm{T}}$. The resulting fuzzy coefficients are depicted in Figure 4.

6.3. Generalization of the example.

Now we extend the example from the previous section to the case when also the output variable $\widetilde{y}$ is fuzzy.

Model I. Let $\widetilde{y}_i$ be the triangular fuzzy number $(y_i - 5k, y_i, y_i + 5k)$, where y_i are taken from Table II and $k = 0, 1, \ldots, 4$. The resulting regression coefficients $\widetilde{b}$, for both relative and absolute tolerances, are depicted in Figure 5. We can see that the more uncertainty in y, the wider are the estimated coefficients $\widetilde{b}$. This illustrates that the width of the regression coefficients $\widetilde{b}$ can be understood as a measure of uncertainty in data.

Model II. Now we consider $\widetilde{y}_i$ to be the Gaussian fuzzy number constructed from the triplet $(y_i - 5k, y_i, y_i + 5k)$ in the same manner as in (14), for $k = 0, \ldots, 4$. The resulting regression coefficients are depicted in Figure 6.

6.4. Conclusions.

We have adapted the tolerance approach for possibilistic linear regression with fuzzy-valued inputs and/or outputs. The method is applicable to any

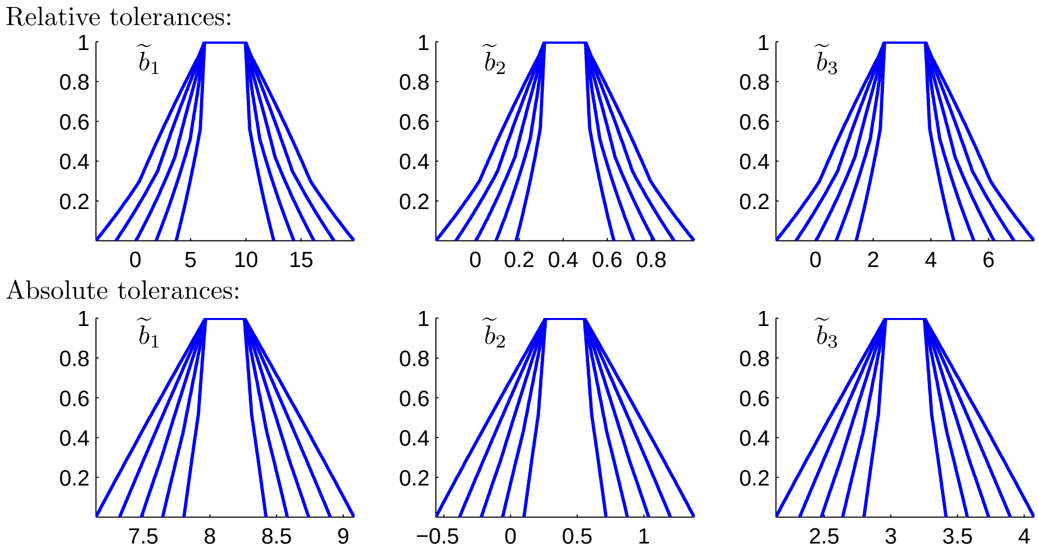

Figure 5: Fuzzy regression coefficients $\widetilde{b}$ for model (13) with triangular inputs $\widetilde{\pi}_i$ and triangular outputs $\widetilde{y}_i = (y_i - 5k, y_i, y_i + 5k)$ with $k = 0, 1, \ldots, 4$.

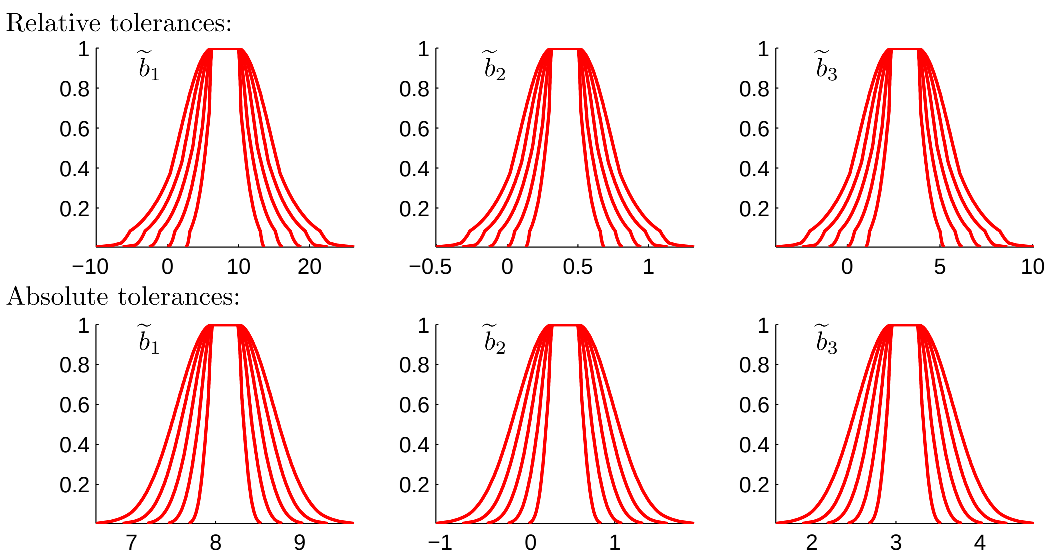

Figure 6: Model II: Gaussian inputs $\widetilde{\pi}_i$ and Gaussian outputs with $k = 0, 1, \ldots, 4$.

class of unimodal fuzzy numbers, not necessarily with a bounded support: in illustrative examples we used both triangular fuzzy data (which are bounded) and asymmetric Gaussian fuzzy data (which are unbounded). The method constructs fuzzy regression coefficients $\widetilde{b}$ respecting the central tendency of a crisp-data estimator applied to defuzzified data, and is minimal with respect to a user-given tolerance vector c. If the data are piecewise linear fuzzy numbers, then the resulting coefficients are piecewise-linear (in the crisp-input-fuzzy-output model) ot piecewise-hyperbolic (in the fuzzy-input-fuzzy-output model). Moreover, the method is computationally very "cheap", and thus can be used for large datasets.

The method constructs reductions *fuzzy model* $\to$ *interval model* $\to$ *crisp model*. With the basic idea in mind, it is straightforward to go further to type-k fuzzy numbers (if a reader considers it to be useful): *type-k fuzzy model* $\to$ *type-*$(k-1)$ *fuzzy model* $\to \cdots \to$ *type-1 fuzzy model* $\to$ *interval model* $\to$ *crisp model*.

Acknowledgment

The work was supported by the Czech Science Foundation under Grant 16-00408S.

References

[1] G. Alefeld and J. Herzberger. *Introduction to interval computations*. Computer Science and Applied Mathematics. Academic Press, New York, 1983.

[2] A. Azadeh, O. Seraj, S. Asadzadeh, and M. Saberi. An integrated fuzzy regression-data envelopment analysis algorithm for optimum oil consumption estimation with ambiguous data. *Appl. Soft Comput.*, 12(8):2614–2630, 2012.

[3] A. Bisserier, R. Boukezzoula, and S. Galichet. Linear fuzzy regression using trapezoidal fuzzy intervals. In B. Bouchon-Meunier, L. Magdalena, M. Ojeda-Aciego, J.-L. Verdegay, and R. Yager, editors, *Foundations of Reasoning under Uncertainty*, volume 249 of *Studies in Fuzziness and Soft Computing*, pages 1–22. Springer Berlin Heidelberg, 2010.

[4] C. Chuang. Extended support vector interval regression networks for interval input-output data. *Inf. Sci.*, 178(3):871–891, 2008.

[5] T. Dereli and A. Durmuşoğlu. Application of possibilistic fuzzy regression for technology watch. *J. Intell. Fuzzy Syst.*, 21(5):353–363, 2010.

[6] C. Filippi. A fresh view on the tolerance approach to sensitivity analysis in linear programming. *Eur. J. Oper. Res.*, 167(1):1–19, 2005.

[7] H. Ge, F. lai Chung, and S. Wang. Theoretical choice of the optimal threshold for possibilistic linear model with noisy input. *IEEE Trans. Fuzzy Syst.*, 16(4):1027–1037, 2008.

[8] P.-Y. Hao. Interval regression analysis using support vector networks. *Fuzzy Sets Syst.*, 160(17):2466–2485, 2009.

[9] M. Hladík. Tolerance analysis in linear systems and linear programming. *Optim. Methods Softw.*, 26(3):381–396, 2011.

[10] M. Hladík and M. Černý. Interval regression by tolerance analysis approach. *Fuzzy Sets Syst.*, 193:85–107, 2012.

[11] D. H. Hong and C. Hwang. Interval regression analysis using quadratic loss support vector machine. *IEEE Trans. Fuzzy Syst.*, 13(2):229–237, 2005.

[12] C.-H. Huang and H.-Y. Kao. Interval regression analysis with soft-margin reduced support vector machine. *Lecture Notes in Computer Science*, 5579 LNAI:826–835, 2009.

[13] C. Hwang, D. H. Hong, and K. Ha Seok. Support vector interval regression machine for crisp input and output data. *Fuzzy Sets Syst.*, 157(8):1114–1125, 2006.

[14] H. Ishibuchi and H. Tanaka. Several formulations of interval regression analysis. In *Proceedings of Sino-Japan Joint Meeting on Fuzzy Sets and Systems*, pages (B2–2)1–4, Beijing, China, 1990.

[15] J.-T. Jeng, C.-C. Chuang, and S.-F. Su. Support vector interval regression networks for interval regression analysis. *Fuzzy Sets Syst.*, 138(2):283–300, 2003.

[16] H. Lee and H. Tanaka. Fuzzy regression analysis by quadratic programming reflecting central tendency. *Behaviormetrika*, 25(1):65–80, 1998.

[17] H. Lee and H. Tanaka. Upper and lower approximation models in interval regression using regression quantile techniques. *Eur. J. Oper. Res.*, 116(3):653–666, 1999.

[18] C. Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

[19] A. Neumaier. *Interval methods for systems of equations*. Cambridge University Press, Cambridge, 1990.

[20] H. Tanaka. Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets Syst.*, 24(3):363–375, 1987.

[21] H. Tanaka, I. Hayashi, and J. Watada. Possibilistic linear regression analysis for fuzzy data. *Eur. J. Oper. Res.*, 40(3):389–396, 1989.

[22] H. Tanaka and H. Lee. Fuzzy linear regression combining central tendency and possibilistic properties. In *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, volume 1, pages 63 –68, Barcelona, Spain, 1997.

[23] H. Tanaka and H. Lee. Interval regression analysis by quadratic programming approach. *IEEE Trans. Fuzzy Syst.*, 6(4):473–481, 1998.

[24] H. Tanaka, S. Uejima, and K. Asai. Linear regression analysis with fuzzy model. *IEEE Trans. Syst., Man Cybern.*, SMC-12(6):903–907, 1982.

[25] H. Tanaka and J. Watada. Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets Syst.*, 27(3):275–289, 1988.