On the Tolerance Approach to Possibilistic Nonlinear Regression over Interval Data

Milan Hladík and Michal Černý

First author: Charles University, Faculty of Mathematics and Physics, Department of Applied Mathematics, Malostranské nám. 25, 11800, Prague, Czech Republic; Department of Econometrics, University of Economics, Prague; Náměstí Winstona Churchilla 4, CZ13067 Prague 3, Czech Republic, `milan.hladik@matfyz.cz`.

Second author: Department of Econometrics, University of Economics, Prague; Náměstí Winstona Churchilla 4, CZ13067 Prague 3, Czech Republic, `cernym@vse.cz`.

Abstract. We study the tolerance-based approach to possibilistic nonlinear regression models with interval data. We provide a method for determination of interval regression parameters of the model for the crisp input – interval output case and for the interval input – interval output case. We define two classes of nonlinear regression models for which efficient algorithms exist. We illustrate the theory by examples.

Keywords: interval regression, nonlinear regression, possibilistic regression, tolerance quotient

1. Introduction

Nonlinear regression is a fundamental tool in data analysis. In this text we address the tolerance approach to possibilistic nonlinear regression, which is a natural generalization of the same concept used in linear regression (Hladík and Černý, 2010; Hladík and Černý, 2011). Possibilistic interval regression was pioneered by (Tanaka, Uejima and Asai, 1987; Tanaka, 1987) in the field of linear regression, and later extended to nonlinear regression (Hao, 2009; Hwang, Hong and Seok, 2006; Jen, Chuang and Su, 2003; Lingras and Butz, 2011; Xu, Luo, Xu and Zhang, 2009), mostly by means of support vector machine. Possibilistic regression was successfully applied in economic forecasting (Lin, Hung and Wu, 2011), system identification (Kaneyoshi, Tanaka, Kamei and Furuta, 1990), speech learning systems (Liu, 2009), or analytic hierarchy process (Entani and Inuiguchi, 2010), among others.

In this text we propose a very general framework for classification of nonlinear regression models which allows us to construct algorithms for computing their possibilistic interval regression parameters. This is useful in particular in case when data to be modeled are of interval nature.

The paper is organized as follows. First we review the notion of possibilistic regression, used in linear regression, and provide a formulation suitable for nonlinear regression models with both crisp input and crisp output data. Then we continue to models involving interval data; in particular, we distinguish crisp input – interval output models and interval input – interval output models. In Section 1.1 we review some examples of nonlinear regression functions widely used applications and in Section 2 we provide a certain general classification framework for nonlinear regression functions. Finally, in Section 3, we state the main problem and design the tolerance-based procedure for computation of interval regression parameters for the classes of functions defined in Section 2.

5th International Conference on Reliable Engineering Computing (REC 2012)
Edited by M. Vořechovský, V. Sadílek, S. Seitl, V. Veselý, R. L. Muhanna and R. L. Mullen

ISBN 978-80-214-4507-9. Published by Ing. Vladislav Pokorný − LITERA

In this text, a *regression function* is simply a continuous function

$$f(x, \theta) = f(x_1, \ldots, x_m; \theta_1, \ldots, \theta_p),$$

where $x_1, \ldots, x_m$ are *data variables* and $\theta_1, \ldots, \theta_p$ are *parameters*.

Given a dataset of n observations of the form

$$(y_i; x_{i*}) = (y_i; x_{i1}, \ldots, x_{im}), \quad i = 1, \ldots, n,$$

the *possibilistic regression* seeks for interval parameters

$$[\underline{\theta}, \overline{\theta}] = [\underline{\theta_1}, \overline{\theta_1}], \ldots, [\underline{\theta_p}, \overline{\theta_p}] \tag{1}$$

such that

$$\forall i \in \{1, \ldots, n\} \; \exists \theta \in [\underline{\theta}, \overline{\theta}] \text{ s.t. } f(x_{i*}, \theta) = y_i.$$

If the condition is satisfied for a given i, we say that *i-th observation is covered*. (Outside data analysis, the problem is sometimes referred to simply as "covering problem" or "envelopment problem".)

Of course, the problem of "finding the interval parameters (1)" must be stated more precisely. Usually we want the find the intervals *as narrow as possible* such that all observations are covered. But that is a multi-criteria optimization problem. The *tolerance approach* is a natural (but not the only possible) method of conversion of the multi-criteria problem to a single-criterion problem. Details of the approach will be discussed in Section 3. As shown in (Hladík and Černý, 2011) (where possibilistic linear regression is studied), the approach has several interesting theoretical properties.

Before we turn into theory, we review some examples of nonlinear regression functions useful in various fields of science and engineering.

1.1. Examples of useful nonlinear regression functions

We sketch only a few examples; more on applications of nonlinear models can be found in (Ratkowski, 1988; Seber and Wild, 2003).

Example. Nonlinear regression functions are often solutions to differential equations describing processes in physics, chemistry or biology. An interesting example is the class of growth curves describing the growth of populations. The Richard's Growth Equation ((Seber and Wild, 2003), p. 332) has a solution

$$f(x; \theta_1, \theta_2, \theta_3, \theta_4) = \theta_1 \cdot (1 + (\theta_4 - 1)e^{-\theta_2(x-\theta_3)})^{1/(1-\theta_4)},$$

which is known as the Richard's Curve. This model has interesting special cases: setting $\theta_4 = 2$ we get the logistic curve

$$f(x; \theta_1, \theta_2, \theta_3) = \frac{\theta_1}{1 + e^{-\theta_2(x-\theta_3)}}, \tag{2}$$

the limit case $\theta_4 \to 1$ yields the Gompertz Curve

$$f(x; \theta_1, \theta_2, \theta_3) = \theta_1 \cdot e^{-e^{-\theta_2(x-\theta_3)}} \tag{3}$$

and the special case with $\theta_4 = 0$ is the model of growth with exponential slow-down

$$f(x; \theta_1, \theta_2, \theta_3) = \theta_1 \cdot (1 - e^{-\theta_2(x-\theta_3)}). \tag{4}$$

Example. The submodel of the Logistic Model (2)

$$f(x;\theta_2,\theta_3) = \frac{1}{1+e^{-\theta_2(x-\theta_3)}}, \quad [\theta_1 \equiv 1] \tag{5}$$

(and similarly with other growth models) is often used when y is interpreted as the probability of an event, where the probability grows with x. For example, we can consider $x =$ pressure and $y =$ probability that the device or material under investigation will be damaged by the pressure. Another example: y can measure the response of a patient to the quantity of drug x.

Example. Another interesting example is the problem of estimation of the degree of polynomial:

$$f(x;\theta_1,\theta_2,\theta_3) = \theta_1 + \theta_2 x + \theta_3 x^{\theta_4}. \tag{6}$$

Example. Berry's Model (Berry, 1967) describes the crop yield as a function of density of planting (or, equivalently, the area available to each plant). Let x_1 be the distance between plants in a row and x_2 the distance between rows of plants. Berry used a model of the form

$$f(x;\theta_1,\theta_2,\theta_3,\theta_4) = \left(\theta_1 + \theta_2\left(\frac{1}{x_1}+\frac{1}{x_2}\right) + \frac{\theta_3}{x_1 x_2}\right)^{-\theta_4}. \tag{7}$$

Example. In physics, the simple oscillation model is important:

$$y = \theta_1 e^{-\theta_2 x}\cos(\theta_3 x). \tag{8}$$

Example. An important class of nonlinear models is the class of models involving a structural change. The basic example is continuous connection of two lines:

$$f(x;\theta_1,\theta_2,\theta_3,\theta_4) = \begin{cases} \theta_1 + \theta_2 x & \text{for } x \le \theta_4, \\ \theta_1 + \theta_4(\theta_2 - \theta_3) + \theta_3 x & \text{for } x > \theta_4. \end{cases} \tag{9}$$

1.2. Notation

Let A^* denote the closure of a set $A \subseteq \mathbb{R}^n$. Given an interval $\boldsymbol{a} \subseteq \mathbb{R} \cup \{\pm\infty\}$, the numbers $\underline{a}$ and $\overline{a}$ denote its lower and upper boundary points, respectively, and a^c and a^Δ denote its center and radius, respectively. That is, $\boldsymbol{a}^* = [\underline{a}, \overline{a}] = [a^c - a^\Delta, a^c + a^\Delta]$. Given a function f and a set A, the symbol $f(A)$ denotes the image of A under f. In particular, $f(\boldsymbol{a})$ stands for the image of an interval $\boldsymbol{a}$.

2. Classes of Nonlinear Regression Models

In order to solve interval nonlinear regression problems, we have to know how to compute image of a function over intervals. Formally, we consider a class of functions equipped by algorithms for determining their images.

Definition 1. *Let*

$$(f_1, f_1^L, f_1^U), \ldots, (f_K, f_K^L, f_K^U) \tag{10}$$

be a set of triples, where for all $k = 1, \ldots, K$:

- $f_k : \mathbb{R} \to \mathbb{R}$ *is a continuous function,*
- $f_k^L(\underline{x}, \overline{x})$ *is an algorithm computing* $\underline{f_k([\underline{x}, \overline{x}])^*}$*,*
- $f_k^U(\underline{x}, \overline{x})$ *is an algorithm computing* $\overline{f_k([\underline{x}, \overline{x}])^*}$*.*

(a) The set (10) *is called* ***basis****.*

(b) The smallest class of functions (of any number of variables)

- *containing constants and the identity function,*
- *containing the functions* $f_1, \ldots, f_k$ *and* $+, -, \times, \div$ *and*
- *being closed under composition and restriction of domain*

is called ***functional universum*** *and is denoted as* $\mathcal{U}$*.*

Determining the image of a function is a fundamental problem of interval analysis (Moore, Kearfott and Cloud, 2009) and by far not trivial. Indeed, only for certain functions we can do it efficiently.

ARITHMETIC EXPRESSIONS

Interval arithmetic is defined naturally as an image of values over interval domains (Moore, Kearfott and Cloud, 2009). Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be real intervals. Then

$$\begin{aligned}
\boldsymbol{a} + \boldsymbol{b} &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}],\\
\boldsymbol{a} - \boldsymbol{b} &= [\underline{a} - \overline{b}, \overline{a} - \underline{b}],\\
\boldsymbol{a} \cdot \boldsymbol{b} &= [\min(\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{ab}), \max(\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{ab})],\\
\boldsymbol{a} \div \boldsymbol{b} &= [\min(\underline{a} \div \underline{b}, \underline{a} \div \overline{b}, \overline{a} \div \underline{b}, \overline{a} \div \overline{b}), \max(\underline{a} \div \underline{b}, \underline{a} \div \overline{b}, \overline{a} \div \underline{b}, \overline{a} \div \overline{b})].
\end{aligned}$$

Given an aritmetic expression $\mathcal{E}$ for a function f, we can evaluate $\mathcal{E}$ by using interval arithmetic. As long as each interval parameter appears at most once in $\mathcal{E}$, then the result equals the image of f. Otherwise, we obtain only an enclosure (a superset) of the image. For example, consider the function

$$f(x, y) = xy - 2x$$

with $x \in [1, 2]$ and $y \in [3, 4]$. Evaluating by interval arithemtic leads to the enclosure

$$f(\boldsymbol{x}, \boldsymbol{y}) \subseteq [1, 2][3, 4] - 2[1, 2] = [-3, 6].$$

However, f can be expressed in other ways. In the form

$$f(x, y) = x(y - 2)$$

each parameter appears just once, so the interval evaluation is exact, i.e.

$$f(\boldsymbol{x}, \boldsymbol{y}) = [1, 2]([3, 4] - 2) = [1, 4].$$

BASIC FUNCTIONS

For some basic function, computing their images over intervals is a simple task. For instance, $\exp(\boldsymbol{x}) = [\exp(\underline{x}), \exp(\overline{x})]$ as the exponential is increasing. Similarly for the functions $\log, \arctan, \ldots$ Some non-monotone functions are tractable, too, e.g. $\sin, \cos, x^n, \ldots$ Polynomials, however, are hard to evaluate exactly in general.

MONOTONICITY

The assumption that each interval parameter should appear at most once in a given expression is restrictive. Moreover, f may be expressed by other basic functions and operations than $+, -, \times, \div$. A strong tool in such a case is to utilize monotonicity. If $f(x) = f(x_1, \ldots, x_m)$ is monotone with respect to the kth parameter x_k, then we are able to get rid of one interval domain. Provided $f(x)$ is non-decreasing at x_k, $\overline{f(\boldsymbol{x})}$ is attained at $\overline{x}_k$, and $\underline{f(\boldsymbol{x})}$ is attained at $\underline{x}_k$. Similarly for the non-increasing case. In this way, the problem of determining $f(\boldsymbol{x})$ is reduced to the problem of determining $\overline{f(\boldsymbol{x})}$ and $\underline{f(\boldsymbol{x})}$ with smaller number of intervals. Hopefully, the sub-problems are of the previous types so that we can calculate the exact values.

For example, let

$$f(x, y) = \frac{x^2 + 6 - y}{y^2}$$

with $x \in [-1, 2]$ and $y \in [1, 2]$. The function is decreasing with respect to y on the interval domains, so in order to compute the lower limit $\underline{f(\boldsymbol{x}, \boldsymbol{y})}$ we fix $y = \overline{y}$, and calculate

$$f(\boldsymbol{x}, \overline{y}) = \frac{\boldsymbol{x}^2 + 5 - \overline{y}}{\overline{y}^2} = \frac{[-1, 2]^2 + 6 - 2}{2^2} = [1, 2].$$

Analogously, to compute the upper limit $\overline{f(\boldsymbol{x}, \boldsymbol{y})}$ we fix $y = \underline{y}$, and calculate

$$f(\boldsymbol{x}, \underline{y}) = \frac{\boldsymbol{x}^2 + 6 - \underline{y}}{\underline{y}^2} = \frac{[-1, 2]^2 + 6 - 1}{1^2} = [5, 9].$$

Putting together, we conclude $f(\boldsymbol{x}, \boldsymbol{y}) = [1, 9]$.

2.1. CLASS OF SUITABLE FUNCTIONS

From the above considerations it is clear that for a well-defined class of function we can determine their images over intervals effectively. For the purpose of interval nonlinear regression, we define the following classes.

Definition 2. *We define the classes of functions $\mathcal{A}$ and $\mathcal{B}$ as follows. Let*

$$f(x, \theta) = f(x_1, \ldots, x_m; \theta_1, \ldots, \theta_p) \in \mathcal{U}.$$

The function $f(x, \theta)$ belongs to the class $\mathcal{A}$ if the function can be analytically expressed such that

– *the expression consists of operations* $+, -, \times, \div$ *and basic functions that are easy to evaluate over intervals, and*

– f *is monotone with respect to the parameters* $\theta_1, \ldots, \theta_p$ *that appear more than once in the expression.*

The function $f(x, \theta)$ *belongs to the class* $\mathcal{B}$ *if the function can be analytically expressed such that*

– *the expression consists of operations* $+, -, \times, \div$ *and basic functions that are easy to evaluate over intervals, and*

– f *is monotone with respect to* $\theta_1, \ldots, \theta_p$, $x_1, \ldots, x_m$ *that appear more than once in the expression.*

We say that the function f *is of* ***type*** $\mathcal{A}$ *and* $\mathcal{B}$*, respectively.*

The significance of the Definition will be clarified in Section 3.1. If a nonlinear regression model is of type $\mathcal{A}$, then there is an efficient method for the tolerance-based possibilistic regression in the crisp input – interval output model. Observe that the crisp input – crisp output model is a special case, hence we also get an algorithmic method for this case as well. If a nonlinear regression model is of type $\mathcal{B}$, then there is an efficient method for the tolerance-based possibilistic regression in the interval input – interval output model.

2.2. EXAMPLES

Whenever we find out that a particular nonlinear regression function belongs to some of the classes $\mathcal{A}, \mathcal{B}$, we know that the nonlinear tolerance approach can be applied to it.

Consider the basis (10) containing $\exp$ and $\ln$. In both cases, the corresponding algorithms f^L and f^U are trivial.

Example. The growth curves (2), (3) and (4) are $\mathcal{B}$-functions.

Example. The regression function (6) can be written in the form

$$y = \theta_1 + \theta_2 x + \theta_3 e^{\theta_4 \ln x},$$

and hence it is an $\mathcal{A}$-type function. If we admit the logarithmic transformation of data $x' := \ln x$, we arrive at the form

$$y = \theta_1 + \theta_2 e^{x'} + \theta_3 e^{\theta_4 x'},$$

and in this form it is an $\mathcal{A}$-type function even if we do not have $\ln$ in the basis. But note that in general, the results of the tolerance-based approach procedure of estimation of interval regression parameters is invariant neither under reparametrization of the model nor under data transformations.

Example. This example shows that a suitable reparametrization of a nonlinear regression function might improve its classification. The Logistic Function is often written in the form

$$f(x; \theta_2, \theta_3) = \frac{e^{\theta_2(x-\theta_3)}}{1 + e^{\theta_2(x-\theta_3)}},$$

where both the variable x and the parameters θ_2, θ_3 occur twice, and hence in this form it is not an $\mathcal{A}$-type function; but its equivalent form (5) is an $\mathcal{B}$-type function.

Example. Berry's Model (7) is an $\mathcal{A}$-function.

Example. The model (9) is an $\mathcal{A}$-type function e. g. under the restriction $\theta_1\theta_3 \geq 0$.

3. The possibilistic interval nonlinear regression

3.1. Setting the problem

Here, we formulate the interval regression problem. Tha aim is to find interval domains for parameters such that all observations are covered by some realization of intervals:

> Find the minimal interval domains for parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)$ such that for every $i = 1, \dots, n$ one has

$$\boldsymbol{y}_i \subseteq f(\boldsymbol{x}_{i*}, \boldsymbol{\theta}). \tag{11}$$

This formulation covers also problems with crisp input or crisp output as special cases. The minimality means that there is no other interval vector $\boldsymbol{\theta}' \subsetneq \boldsymbol{\theta}$ satisfying (11). Nevertheless, there may exist other interval vector, or typically many of them, that is also minimal with respect to inclusion. So there are many degrees of freedom which minimal solution to consider. To obtain good interval parameters, the following properties should be more or less satisfied:

- The radii of interval parameters, $\theta_1^\Delta, \dots, \theta_p^\Delta$ are balanced. It is undesirable when some interval is very narrow, or even crisp, while another is very wide.
- The interval parameters follow the so called *central tendency*. That is, their centers more or less fit the data with respect to traditional goodness-of-fit measures.
- The method is not much sensitive to outliers.

In order to fulfill these requirements for interval linear regression models, the authors proposed in (Hladík and Černý, 2010; Hladík and Černý, 2011) a two level method. In the first step, we calculate crisp estimation $\theta^c = (\theta_1^c, \dots, \theta_p^c)$ to the nonlinear regression model. In the second step, we minimally extend the parameters to intervals such that they cover all observations. This basic idea is usable for nonlinear regression as well; we do it in the next section.

3.2. Methodology

As indicated in the previous section, we calculate interval parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)$ in two steps:

(a) Compute the centers $\theta^c = (\theta_1^c, \dots, \theta_p^c)$;

(b) Compute the radii $\theta^\Delta = (\theta_1^\Delta, \dots, \theta_p^\Delta)$.

Centers are determined by any traditional method for nonlinear regression. In case of interval input or output, we take the centers of the intervals. Thus, we have a standard nonlinear regression model with crisp data and can apply any method to compute θ^c. This makes our approach flexible, since it doesn't rely on one concrete algorithm. Next, the property on central tendency is fulfilled, too.

In the second step, we calculate the radii θ^Δ. In order that the resulting interval parameters are balanced with respect to their radii, we introduce *tolerance rates* as a non-negative vector $c^\Delta = (c_1^\Delta, \dots, c_p^\Delta)$. The radii of interval parameters are then considered in the form $\theta^\Delta = \delta c^\Delta$, or

$$(\theta_1^\Delta, \dots, \theta_p^\Delta) = (\delta c_1^\Delta, \dots, \delta c_p^\Delta),$$

where $\delta > 0$ is the unknown *tolerance quotient*. The aim is to determine the minimal tolerance quotient such that the corresponding interval parameters cover all observations. A tolerance quotient satisfying the coverage condition is called feasible.

The tolerance rates are usually set up as $c^\Delta = |\theta^c|$ or $c^\Delta = (1, \dots, 1)$. The former corresponds to relative perturbations, while the latter force all interval parameters to have the same width. If the kth interval parameter is desired to be crisp, so it suffices to put $c_k^\Delta = 0$.

Now, all we need is to compute the minimal feasible tolerance quotient $\delta > 0$. We employ the bisection method. Denote $\boldsymbol{\theta}^\delta := [\theta^c - \delta c^\Delta, \theta^c + \delta c^\Delta]$ the form of the resulting interval parameters. The basic algorithmic scheme is as follows:

1. Put $\delta = 1$ and loop the following command for a given number of iterations.
2. If $\boldsymbol{y}_i \subseteq f(\boldsymbol{x}_{i*}, \boldsymbol{\theta}^\delta)$ for every $i = 1, \dots, n$, then decrease δ. Otherwise, increase δ.

Denote by δ^* the return value of δ. Notice that provided the amount of decrease and increase of δ is halved, the iterations converge exponentially fast to the optimum. Thus, for practical purposes, 5 to 15 iterations are usually enough to provide us with a sufficiently accurate approximation. For a model of type $\mathcal{A}$ or $\mathcal{B}$, the evaluation of the image $f(\boldsymbol{x}_{i*}, \boldsymbol{\theta}^\delta)$ is fast, therefore, the overall time complexity of the algorithm is mild.

If the last iteration was the decrease of δ, we increase δ^* correspondingly in order to obtain a feasible δ^*. However, it may still happen that δ^* is not feasible. We indicate it easily be observing that δ was never decreased in the run of the algorithm. This situation happens rarely, but cannot be excluded. For example, consider the Gompertz Curve of the form

$$f(x, \theta_1) = e^{-e^{x-\theta_1}}.$$

It is easily seen that for any θ_1, $f((-\infty, \infty), \theta_1) = (0, 1)$. If our dataset contains, say, a point $(x = 0, y = 1)$, that point cannot be covered, which implies that the algorithm tends to increase δ up to infinity. In general, from the algorithmic point of view, the problem whether a given point can be covered (with a possibly huge value of δ), is undecidable; hence we cannot do anything else than terminating the algorithm when the value of δ exceeds limits in which the value δ has reasonable interpretation for the regression model under consideration.

3.3. Properties of the model

Here we only sketch some properties of the model, which have been investigated in (Hladík and Černý, 2011) in the case of linear regression models.

- The method is flexible by utilizing any traditional method for the parameteric centers.
- If the input model is of type $\mathcal{A}$ or $\mathcal{B}$, then the optimal tolerance quotient, and thus the minimal interval parameters, are computed efficiently with a given precision.
- The interval parameters have balanced widths, proportional to the apriori given rates.
- Outliers can be handled.

Concerning outliers, they can be managed in many ways, depending on the purposes of decision maker. For instance, the method is easily adapted to the model, in which only a fraction, say 90%, of observations should be covered. Another possibility is to calculate the tolerance quotient δ^* such that the corresponding interval parameters cover e.g. 80% observations, and then consider as outliers all observations that are not covered by the tolerance quotient $1.1\delta^*$.

3.4. Examples

Example 1. Assume that we measure reliability of a material (y) as a function of time (x) for which the material is exposed to unfavorable conditions (such as unfavourable temperature or pressure). Of course it can be expected that the longer the exposition is, the higher level of disruption. Assume that the level of disruption is measured on a discrete scale $0, \ldots, 10$, where 0 means "no damage", 1 means "very mild damage", $\ldots$, and 10 means "totally damaged". Assume further that the values of y are determined by experts (say, by visual inspection of constructions where the material has been used). Due to a certain subjectivity of experts, it is appropriate to consider the grade $y \in \{1, \ldots, 9\}$ as an interval, say of the form

$$[\underline{y}, \overline{y}] = [y - 0.5, y + 0.5] \tag{12}$$

We model the dependence of y on x using the Gompertz curve

$$y = 10e^{-e^{-\theta_1(x-\theta_2)}}, \tag{13}$$

where θ_1 measures slope of the curve (that is, the speed of worsening of the condition of the material) and θ_2 measures the shift of the curve. The shift measures whether the process of wearing of the material starts earlier or later.

Assume that we have data from Table I. Using nonlinear least squares on the data $(x_1, y_1), \ldots, (x_{30}, y_{30})$, we fit

$$\widehat{\theta}_1 = 0.795, \quad \widehat{\theta}_2 = 4.887. \tag{14}$$

This curve describes "average" behavior of the material with respect to x.

Now we would like to extend the estimated crisp values $\theta_1^c = \widehat{\theta}_1$ and $\theta_2^c = \widehat{\theta}_2$ to interval values covering all observations, taking into account the fact that it is more appropriate to handle an observation y as an interval (12) rather than a fixed value.

We observe that the points $y \in \{0, 10\}$ can never be covered with the Gompertz curve. We take the following step. We divide data into three categories:

- A: material is unaffected by the unfavorable conditions;

- B: phase of wearing;
- C: phase of total weariness.

We assume that the phase B starts when the first mild defect is encountered (i.e. the first time with $y \geq 1$) and that the phase C starts when the first total damage is encountered (i.e. first time with $y = 10$). The division of data is also shown in Table I.

The main purpose of the Gompertz curve is modeling the dynamics of the wearing process, which corresponds to the phase B. Hence it makes sense to take into account only B-data and apply the tolerance method to them. (Observe that the data point $(x_{24}, [\underline{y}_{24}, \overline{y}_{24}])$, being a C-point, need not be covered.)

As a first example, we set $c^{\Delta} = \binom{0.795}{4.887}$ (i.e., relative tolerances). We arrive at the value

$$\delta^* = 0.183.$$

Hence we conclude that it suffices to perturb the values $\widehat{\theta}_1, \widehat{\theta}_2$ by no more that 18.3% in order all intervals be covered. We can roughly say that "the truth" is covered by the intervals $[(1-0.183)\cdot 0.795, (1+0.183)\cdot 0.795]$ and $[(1-0.183)\cdot 4.887, (1+0.183)\cdot 4.887]$ for θ_1 and θ_2, respectively. The resulting data enclosure is plotted in Figure 1 with a dotted line. We can also say that the "pessimistic scenario" for the speed of weariness (measured by θ_1) is $(1+0.183)\cdot 0.795 = 0.94$ and that the "pessimistic scenario" for the shift is $(1-0.183)\cdot 4.887 = 3.99$.

As a second example, we set $c^{\Delta} = \binom{1}{1}$ (i.e., absolute tolerances). We arrive at the value

$$\delta^* = 0.360.$$

Now the data are covered by the intervals $[0.795 - 0.36, 0.795 + 0.36]$ and $[4.887 - 0.36, 4.887 + 0.36]$ for θ_1 and θ_2, respectively. The resulting data enclosure is plotted in Figure 1 with a dashed-dotted line.

As a third example, we set $c^{\Delta} = \binom{0}{4.887}$. This models the situation that the speed of worsening is kept constant and we can perturb only the shift θ_2 to cover the data. (Hence we seek for an interval for θ_2 only.) We arrive at the value

$$\delta^* = 0.254.$$

Now the data are covered by the interval $[(1-0.36)\cdot 4.887, (1+0.36)\cdot 4.887]$ for θ_2. The resulting data enclosure is plotted in Figure 1 with a dashed line. Now we can say: if we know that the speed of wearing is $\theta_1 = 0.795$, then the pessimistic scenario for θ_2 is $(1-0.36)\cdot 4.887 = 3.13$.

Example 2. In Example 1 we used the fact that the Gompertz function (13) is $\mathcal{A}$-type function. Using the fact that it is also the $\mathcal{B}$-type function, we can extend the example to the case where x-data are of interval nature. This corresponds to the situation that we do not know exactly the times in which the measurements were made. Again we use the data from Table I and for each of the observations we assume that its x-value is an interval

$$[x_i - \tfrac{1}{2}, x_i + \tfrac{1}{2}].$$

We set the values $(\theta_1^c, \theta_2^c) = (\widehat{\theta}_1, \widehat{\theta}_2)$ from (14). Using the tolerance method for covering the B-phase data, we arrive at the results

- $c^{\Delta} = \binom{0.795}{4.887}$: $\delta^* = 0.250$,
- $c^{\Delta} = \binom{1}{1}$: $\delta^* = 0.61$,

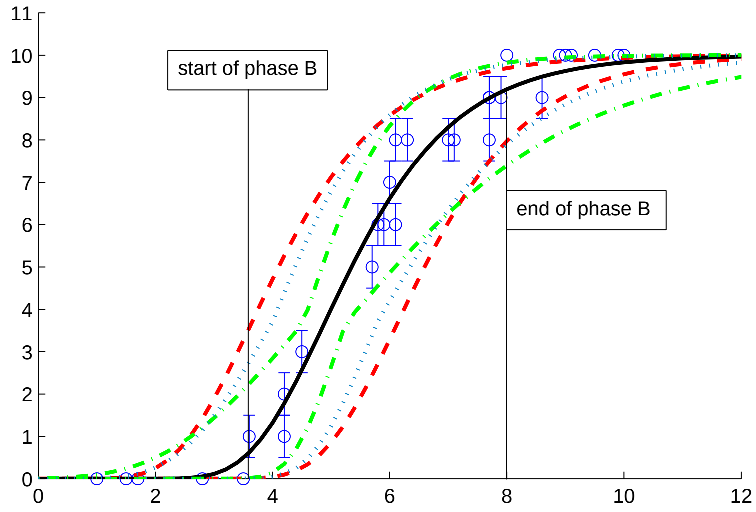

Figure 1. Data from Table I, the crisp model (13) with $\widehat{\theta}_1 = 0.795$ and $\widehat{\theta}_2 = 4.887$ and the enclosures with $c^{\Delta} = (0.79, 4.887)^T$ (relative tolerances, dotted), $c^{\Delta} = (1, 1)^T$ (absolute tolerances, dashed-dotted) and $c^{\Delta} = (0, 4.887)^T$ (only perturbation of θ_2 allowed, dashed).

Table I. Source data for the Example.

phase	i	x_i	y_i	$\underline{y}_i$	$\overline{y}_i$	phase	i	x_i	y_i	$\underline{y}_i$	$\overline{y}_i$
A	1	1.0	0	—	—	B	16	6.3	8	7.5	8.5
A	2	1.5	0	—	—	B	17	7.0	8	7.5	8.5
A	3	1.7	0	—	—	B	18	7.1	8	7.5	8.5
A	4	2.8	0	—	—	B	19	7.7	9	8.5	9.5
A	5	3.5	0	—	—	B	20	7.7	8	7.5	8.5
B	6	3.6	1	0.5	1.5	B	21	7.7	9	8.5	9.5
B	7	4.2	2	1.5	2.5	B	22	7.9	9	8.5	9.5
B	8	4.2	1	0.5	1.5	C	23	8.0	10	—	—
B	9	4.5	3	2.5	3.5	C	24	8.6	9	8.5	9.5
B	10	5.7	5	4.5	5.5	C	25	8.9	10	—	—
B	11	5.8	6	5.5	6.5	C	26	9.0	10	—	—
B	12	5.9	6	5.5	6.5	C	27	9.1	10	—	—
B	13	6.0	7	6.5	7.5	C	28	9.5	10	—	—
B	14	6.1	8	7.5	8.5	C	29	9.9	10	—	—
B	15	6.1	6	5.5	6.5	C	30	10.0	10	—	—

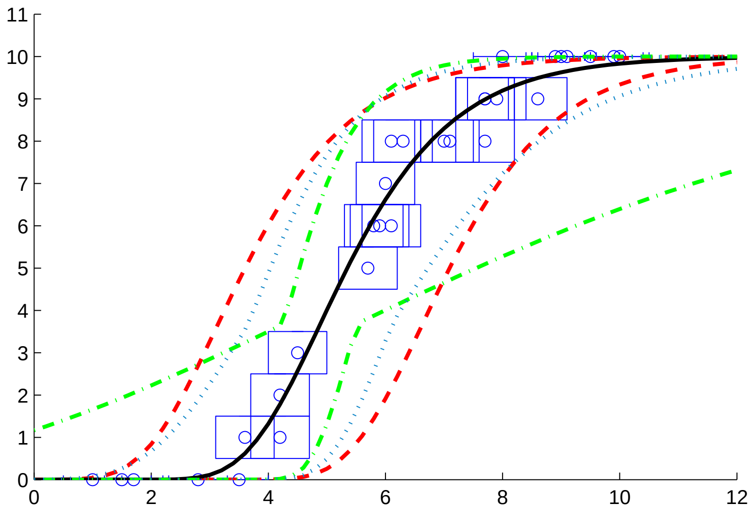

Figure 2. Data from Table I with interval-valued x_i's in the form $\boldsymbol{x}_i = [x_i - \frac{1}{2}, x_i + \frac{1}{2}]$, the crisp model (13) with $\widehat{\theta}_1 = 0.795$ and $\widehat{\theta}_2 = 4.887$ and the enclosures with $c^{\Delta} = (0.79, 4.887)^T$ (relative tolerances, dotted), $c^{\Delta} = (1,1)^T$ (absolute tolerances, dashed-dotted) and $c^{\Delta} = (0, 4.887)^T$ (only perturbation of θ_2 allowed, dashed).

– $c^{\Delta} = \binom{0}{4.887}$: $\delta^* = 0.357$,

with the resulting enclosures depicted in Figure 2. Recall that the data point $([x_{24} - \frac{1}{2}, x_{24} + \frac{1}{2}], [\underline{y}_{24}, \overline{y}_{24}])$, being a C-point, need not be covered.

4. Conclusions

In this text we extended the tolerance-based approach, originally designed for possibilistic linear regression, for a particular class of nonlinear regression models. The method provides a covering of either crisp or interval data of the model and for that class of models it can be computed by an efficient algorithm (provided that the algorithms f^U and f^L for the basic functions are efficient). For the class of non-$\mathcal{A}$-type models, the method provides only lower bound on the optimal tolerance rate δ^* in general. The interesting question for further research is whether and under which conditions the method could be adapted for a wider of nonlinear models to yield the optimal value.

Acknowledgements

The authors were partially supported by the Czech Science Foundation Grant P403/12/1947.

References

Berry, G. A Mathematical Model Relating Plant Yield with Arrangement for Regularly Spaced Crops. *Biometics* 23: 505–515, 1967.

Entani, T. and M. Inuiguchi. Group Decisions in Interval AHP Based on Interval Regression Analysis. In Huynh, V.-N. et al. (editors), *Integrated Uncertainty Management and Applications*, vol. 68 of Advances in Soft Computing, 269–280, Springer, 2010.

Hao, P.-Y. Interval Regression Analysis Using Support Vector Machines. *Fuzzy Sets and Systems* 160 (17): 2466–2485, 2009.

Hladík, M. and M. Černý. New Approach to Interval Linear Regression. In Kasımbeyli et al. (eds.), *24th Mini-EURO Conference on Continuous Optimization and Information-Based Technologies in the Financial Sector MEC EurOPT 2010. Selected Papers,* 197–171, Technika, Vilnius, 2010.

Hladík, M. and M. Černý. Interval Regression by Tolerance Analysis Approach. *Fuzzy Sets and Systems* (in press), 2011, DOI 10.1016/j.fss.2011.07.009.

Hwang, C., D. H. Hong and K. Ha Seok. Support Vector Interval Regression Machine for Crisp Input and Output Data. *Fuzzy Sets and Systems* 157 (8): 1114–1125, 2006.

Jen J.-T., C.-C. Chuang and S.-F. Su. Support Vector Interval Regression Networks for Interval Regression Analysis. *Fuzzy Sets and Systems* 138 (2): 283–300, 2003.

Kaneyoshi, M., H. Tanaka, M. Kamei and H. Furuta. New System Identification Technique Using Fuzzy Regression Analysis. In *Proceedings of the First International Symposium on Uncertainty Modeling and Analysis*, 258–533, Baltimore, USA, 1990.

Lin, K.-P., K.-C. Hung and M.-C. Wu. Long-term Business Cycle Forecasting Using Intuitionistic Fuzzy Least-squares Suppert Vector Regression. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2495–2499, Taipei, Taiwan, 2011.

Lingras, P. and C. Butz. Conservative and Aggressive Rough SVR Modeling. *Theoretical Computer Science* 412 (42): 5885–5901, 2011.

Liu, P. Study on a Speech Learning Approach Based on Interval Support Vector Regression. In *Proceedings of 41th International Conference on Computer Science & Education*, 1009-1012, Nanning, China, 2009.

Moore, R. E., R. B. Kearfott and M. J. Cloud. *Introduction to Interval Analysis.* SIAM, Philadelphia, 2009.

Ratkowski, D. A. *Handbook of Nonlinear Regression Models.* Marcel Dekker, 1988.

Seber, G. A. F. and C. J. Wild. *Nonlinear Regression.* Wiley, 2003.

Tanaka, H. Fuzzy Data Analysis by Possibilistic Linear Models. *Fuzzy Sets and Systems* 24 (3): 363–375, 1987.

Tanaka, H., S. Uejima and K. Asai. Linear Regression Analysis with Fuzzy Model. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-12 (6): 903–907, 1982.

Xu, S.-Q., Q.-Y. Luo, Xu G.-H. and L. Zhang. Assymetrical Interval Regression using Extended ε-SVM with Robust Algorithm. *Fuzzy Sets and Systems* 160 (7): 988–1002, 2009.