The regression tolerance quotient in data analysis

Michal Černý [1], Milan Hladík [2]

Abstract. Let $E(y) = Xb$ be the traditional linear regression model and let $\widehat{b}$ be an estimate of the unknown vector of regression parameters b. The tolerance quotient δ^*, defined and studied by Hladík and Černý in (*Interval regression by tolerance analysis approach*, preprint in KAM-DIMATIA Series 963, 2010) is the least $\delta \geq 0$ such that for any i, the equation $y_i = X_i\beta$, where y_i is the i-th observation of the dependent variable and X_i is the i-th row of X, is satisfied with some $\beta \in [\widehat{b} - \delta \cdot |\widehat{b}|, \widehat{b} + \delta \cdot |\widehat{b}|]$. The tolerance quotient δ^* measures the relative perturbation rate, i.e. how much it is necessary to perturb the estimated regression coefficients $\widehat{b}$ to satisfy each of the equations $y_i = X_i\beta$, and hence is a measure of goodness of fit of the model. We demonstrate the usage of the quotient in analysis of both crisp and interval data and, in particular, interval data arising in econometrics and finance. We show a method to study probabilistic properties of the tolerance quotient: we derive the distribution of δ^* and, under certain assumptions, we present a method for construction of a confidence interval for δ^*.

Keywords: tolerance quotient, interval regression, confidence interval

JEL classification: C52
AMS classification: 62J86

1 Introduction

The traditional linear regression model $E(y) = Xb$ describes the response of the dependent variable y as a linear function of dependent variables X. The vector b of regression parameters is unknown and it is to be estimated. The most common estimator is OLS: that is, finding $\widehat{b}$ such that the L_2-norm of residuals $r := y - X\widehat{b}$ is minimized. In robust regression, L_1-norm is often used instead of L_2; it is well known that this minimization problem is reducible to linear programming.

The traditional approach assumes that the observed values of independent variables (the rows of the design matrix X) and observations of the dependent variable (the components of the vector y) are *crisp*, i.e. they are real (or rational) numbers. In many practical applications, some or all of the values X and y cannot be directly observed; they might be uncertain or fuzzy. Only an interval, in which the unobservable value is guaranteed to be, is known.[†] In this context it makes sense to generalize the traditional linear regression model to be able to handle *intervals*.

Interval variables appear in economic and financial applications quite often, for example:

- traded variables have bid-ask spread;
- credit rating grades are sometimes regarded as intervals of credit spreads above the risk-free yield curve (though this interpretation is simplified);
- if we measure economic variables such as personal income, we sometimes obtain underestimated observations due to the presence of the 'grey zone'. For example, if personal income Y is measured by means of the income declared, the true income is likely to be in an interval $[Y, Y + \Delta_Y]$ where Δ_Y is an upper bound for 'grey' (undeclared) income;

[1] University of Economics Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech republic, e-mail: cernym@vse.cz

[2] Faculty of Mathematics and Physics, Charles University, Department of Applied Mathematics, Malostranské náměstí 25, CZ11801 Prague, Czech republic, e-mail: hladik@matfyz.cz

[†] This issue also arises in numerical computations if the computation precision is limited; then, we do not know the true value of a given rational number, we only know that the true value is guaranteed to be in a certain interval, the width of which corresponds to the data type used for representation of rational numbers.

- there is a similar problem with demographic data, e.g. true immigration is in an interval $[I, I + \Delta_I]$ where I is observed legal immigration and Δ_I is an (estimated) upper bound on illegal immigration;
- dynamic macroeconometric regression models often include variables such as foreign exchange rates or interest rates, that are not constant within a given period. Usually, the median or the average value is taken as a proxy. However, it might be more appropriate to regard that variable as an interval within which the variable changes over the period;
- individual data are available only in an interval-censored form.

More applications of interval data analysis and related methods are discussed in [2, 5, 6, 7, 9, 14].

Various versions of interval regression models have been proposed and studied, see e.g. [8]:

- *crisp input – interval output model.* The matrix X is crisp (i.e. it is a matrix of real numbers) and y is an interval vector.
- *interval input – crisp output model.* X is an interval matrix and y is a crisp vector.
- *interval input – interval output model.* Both X and y are intervals.

In this context, the traditional regression model is called *crisp input – crisp output model.*

Several concepts of interval regression have been proposed [8, 11, 13]: the best known approaches are the *possibility concept* and the *necessity concept.* They will be discussed in Sect. 2.

We use boldface for interval matrices. If $\underline{X}_{ij} \leq \overline{X}_{ij}$ for all i and j, then the *interval matrix* $\boldsymbol{X} = [\underline{X}, \overline{X}]$ is the set $\{X : (\forall i)(\forall j) X_{ij} \in [\underline{X}_{ij}, \overline{X}_{ij}]\}$, where X_{ij} is the (i, j)-th component of the matrix X. The interval vector $\boldsymbol{y} = [\underline{y}, \overline{y}]$ is a one-column interval matrix. Its i-th component is denoted $\boldsymbol{y}_i$ $(= [\underline{y}_i, \overline{y}_i])$.

Interval regression models are much more difficult to handle than the traditional linear regression models. Just to illustrate the difficulties, consider one of the most fundamental problems: to describe the set of OLS solutions to the interval regression model. In the most general case of the interval input – interval output, the *set of OLS solutions* is $B := \{b : X^{\mathrm{T}}Xb = X^{\mathrm{T}}y,\ X \in \boldsymbol{X},\ y \in \boldsymbol{y}\}$. What can be said about this set? Very little is known. In general, this set need not be bounded and need not be convex. If it is bounded, only very rough and computationally expensive coverings of the set are known, see e.g. [10]. But the situation is even worse: deciding whether B is bounded is a *co-NP*-complete problem [10].

In the case of the crisp input – interval output model, i.e. if $\underline{X} = \overline{X}$, the set B is a zonohedron [15]. This might rise hope that it could be possible to describe the set easily. However, the polyhedron might suffer from a quite complex combinatorial structure. It has (in general) $\Theta(n^d)$ vertices, where n is the dimension of the interval vector (cube) $\boldsymbol{y}$ and d is the number of regression parameters. So it is apparent that even for quite small n and d (say, $n = 100$ and $d = 4$) it is computationally very hard to describe the set B by vertex enumeration. And, moreover, for a user of such a model, this description is not very friendly. It would be natural to describe the zonohednon by linear inequalities; unfortunately, no computationally feasible method is known (to the authors' knowledge).

These are only a few difficulties arising from the inclusion of interval data in regression models. So, other approaches for handling interval regression models have been sought. We shall deal with one such approach, known as the *tolerance approach*; for further references see [4, 5].

2 The tolerance quotient

Let X_j denote the j-th row of a matrix X. The absolute value $|X|$ of a matrix X is understood componentwise.

First we deal with the simplest case of the crisp input – crisp output model; then we will generalize the ideas to interval models.

Crisp input – crisp output model. Let y and X be crisp and let $\widehat{b}$ be an estimate of the vector of regression parameters b; for example, take the OLS estimate $\widehat{b} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$. Let n denote the number of observations and d number of regression parameters. The *tolerance quotient* δ^* is the minimal $\delta \geq 0$ such that

$$(\forall j \in \{1, \ldots, n\})(\exists \beta \in [\widehat{b} - \delta \cdot |\widehat{b}|, \widehat{b} - \delta \cdot |\widehat{b}|]) \;\; y_j = X_j \beta. \tag{1}$$

That is, δ^* is the minimal *perturbation rate*: it is sufficient to perturb the estimated regression coefficients $\widehat{\beta}$ by no more that $100\delta^*\%$ to fulfill each equality in the system $y = X\beta$. So, the tolerance quotient is a measure of goodness-of-fit of the model, an alternative measure to the classical statistics such as R^2.

The basic theorem of the tolerance approach to regression has been proved in [3] and [5]. By convention, the value of a fraction with a zero denominator is zero.

Theorem 1. *If there is $j \in \{1, \ldots, n\}$ such that $|X|_j \cdot |\widehat{\beta}| = 0$ and $y_j \neq X_j \widehat{b}$ then no δ satisfies (1). Otherwise*

$$\delta^* = \max_{1 \leq j \leq n} \frac{|y_j - X_j \widehat{b}|}{|X|_j \cdot |\widehat{b}|}. \quad \square \tag{2}$$

Before we turn to interval models, let us discuss some properties and possibilities to use the tolerance quotient in analysis of crisp input – crisp output models.

Computability. Observe that δ^* is very easily computable. Indeed, to evaluate the expression (2) we need linear computation time only.

Goodness-of-fit measure. Let $\widehat{b}_1$ and $\widehat{b}_2$ be two different estimates of b obtained by two different methods (e.g. by OLS and L_1-norm minimization of residuals). The lower of the tolerance quotients $\delta^*_{\widehat{b}_1}$ and $\delta^*_{\widehat{b}_2}$ (given by (2) with $\widehat{b} := \widehat{b}_1$ and $\widehat{b} := \widehat{b}_2$, respectively) is an indication which estimate should be preferred.

Detection of outliers. It is apparent that existence of outliers in data might significantly increase the tolerance quotient. Hence, the j-th point y_j, where j is the argmax of (2), is likely to be an outlier (if outliers are really present). This suggests a simple procedure for removal of outliers: remove those points, the removal of which decreases the tolerance quotient significantly.

Detection of change points. Regression models sometimes suffer from the existence of structural breaks. In econometrics it is a frequent case; for example, a structural break in a dynamic model may appear in the period when the exchange-rate regime changes, or it may appear with some lag. The structural break, or the change point, is the index $\kappa \in \{1, \ldots, n-1\}$ such that

$$E(y_j) = \begin{cases} X_j b_1 & \text{for } j = 1, \ldots, \kappa, \\ X_j b_2 & \text{for } j = \kappa + 1, \ldots, n, \end{cases}$$

where $b_1 \neq b_2$. There are various methods for testing the existence of structural breaks and there are various methods for estimation of κ (regarded as an unknown parameter, see [1]). The tolerance quotient might be an alternative approach for detection of the location of a structural change. Let

$$\delta^*_{k_1:k_2} = \max_{k_1 \leq j \leq k_2} \frac{|y_j - X_j \widehat{b}_{k_1:k_2}|}{|X|_j \cdot |\widehat{b}_{k_1:k_2}|}, \tag{3}$$

where $\widehat{b}_{k_1:k_2}$ is an estimator for the model $E(y_{k_1:k_2}) = X_{k_1:k_2} b$. We have denoted $y_{k_1:k_2} = (y_{k_1}, y_{k_1+1}, \ldots, y_{k_2})^{\mathrm{T}}$ and $X_{k_1:k_2} = (X^{\mathrm{T}}_{k_1}, X^{\mathrm{T}}_{k_1+1}, \ldots, X^{\mathrm{T}}_{k_2})^{\mathrm{T}}$. For example,

$$\widehat{b}_{k_1:k_2} = \begin{cases} (X^{\mathrm{T}}_{k_1:k_2} X_{k_1:k_2})^{-1} X^{\mathrm{T}}_{k_1:k_2} y_{k_1:k_2} & \text{if } X^{\mathrm{T}}_{k_1:k_2} X_{k_1:k_2} \text{ is regular}, \\ \text{undefined} & \text{otherwise}. \end{cases}$$

By convention, if $\widehat{b}_{k_1:k_2}$ is undefined, then $\frac{|y_j - X_j \widehat{b}_{k_1:k_2}|}{|X_j| \cdot |\widehat{b}_{k_1:k_2}|} = \infty$.

If we want to estimate the point of the structural break, we might inspect the values $d_k := \max\{\delta^*_{1:k}, \delta^*_{k+1:n}\}$; the point $\mathrm{argmin}_k d_k$ may be an indicator for the location of the structural break. More in general, the behavior of the series $\delta^*_{1:k}$ and $\delta^*_{k+1:n}$ may give information complementary to the traditional methods for estimation of changepoints.

Crisp input – interval output model. Now assume that $\boldsymbol{y} = [\underline{y}, \overline{y}]$ and that $\widehat{\beta}$ is available. The usual choice is $\widehat{b} = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} \cdot \frac{1}{2}(\underline{y} + \overline{y})$. At least three approaches for the crisp input – interval output model are considered in literature: the *possibilistic concept*

$$(\forall j \in \{1, \ldots, n\})(\forall \upsilon_j \in [\underline{y}_j, \overline{y}_j])(\exists \beta \in [\widehat{b} - \delta \cdot |\widehat{b}|, \widehat{b} + \delta \cdot |\widehat{b}|]) \ \ \upsilon_j = X_j \beta,$$

the *weak possibilistic concept*

$$(\forall j \in \{1, \ldots, n\})(\exists \upsilon_j \in [\underline{y}_j, \overline{y}_j])(\exists \beta \in [\widehat{b} - \delta \cdot |\widehat{b}|, \widehat{b} + \delta \cdot |\widehat{b}|]) \ \ \upsilon_j = X_j \beta,$$

and the *necessity concept*

$$(\forall j \in \{1,\ldots,n\})(\forall \beta \in [\widehat{b} - \delta \cdot |\widehat{b}|, \widehat{b} + \delta \cdot |\widehat{b}|])(\exists \upsilon_j \in [\underline{y}_j, \overline{y}_j]) \;\; \upsilon_j = X_j \beta.$$

Again, we are searching for δ^*, the minimal $\delta \geq 0$ fulfilling the chosen concept. It is interesting that the first two cases are reducible to the crisp input – crisp output model (2).

Theorem 2 ([3])**.** *The tolerance quotient δ^* for the possibilistic concept is given by (2) with*

$$y_j = \begin{cases} \underline{y}_j & \text{if } |\underline{y}_j - X_j \widehat{b}| \geq |\overline{y}_j - X_j \widehat{b}|, \\ \overline{y}_j & \text{otherwise.} \end{cases}$$

The tolerance quotient δ^* for the weak possibilistic concept is given by (2) with

$$y_j = \begin{cases} \underline{y}_j & \text{if } \underline{y}_j > X_j \widehat{b}, \\ X_j \widehat{b} & \text{if } X_j \widehat{b} \in [\underline{y}_j, \overline{y}_j], \\ \overline{y}_j & \text{otherwise.} \end{cases}$$

For the neccessity concept it holds

$$\delta^* = \min_{1 \leq j \leq n} \min \left\{ \frac{\overline{y}_j - X_j \widehat{b}}{|X|_j \cdot |\widehat{b}|}, \frac{X_j \widehat{b} - \underline{y}_j}{|X|_j \cdot |\widehat{b}|} \right\},$$

where, by convention, the value of a fraction with a zero denominator is ∞. □

Interval input – interval output model. In the interval input – interval output models, there are even more solution concepts. It is interesting that many of them are also reducible to the crisp input – crisp output case. We shall not discuss them any more; see [3].

3 The distribution of δ^*

We have demonstrated the usefulness of the tolerance quotient δ^* in the crisp input – crisp output model. On one hand, it is useful in analysis of traditional regression models as discussed in Sect. 2. On the other hand, it is useful for tolerance analysis of interval regression models as optimal tolerance rates of many solution concepts of the interval models, both crisp input – interval output and interval input – interval output, are reducible to the crisp input – crisp output case.

Now we show some probabilistic properties of δ^* in the crisp input – crisp output model.

So far we have treated y_j's as fixed observations. To get some insight, what values of δ^* we shall expect, assume that y_j, $j = 1,\ldots,n$, are independent normal variables with means $X_j b$ and a common standard error $\sigma > 0$. (These are traditional assumptions in regression analysis.) Then, δ^* may be regarded as a random variable. Let $\varphi(\sigma; x) := (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ and $\Phi(\sigma; x) := \int_{-\infty}^{x} \varphi(\sigma; t) \; \mathrm{d}t$. Denote

$$\Phi_j^*(z) := \begin{cases} 0 & \text{if } z \leq 0, \\ 2\Phi(\frac{\sigma}{|X|_j \cdot |b|}; z) - 1 & \text{if } z > 0 \text{ and } |X|_j \cdot |b| > 0, \\ 1 & \text{if } z > 0 \text{ and } |X|_j \cdot |b| = 0 \end{cases} .$$

and let φ_j^* be the derivative of Φ_j^*:

$$\varphi_j^*(z) := \begin{cases} 0 & \text{if } z \leq 0 \text{ or } (z > 0 \text{ and } |X|_j \cdot |b| = 0), \\ 2\varphi(\frac{\sigma}{|X|_j \cdot |b|}; z) & \text{if } z > 0 \text{ and } |X|_j \cdot |b| > 0. \end{cases}$$

Theorem 3. *The density of δ^* is*

$$\varphi_{\delta^*}(z) := \sum_{j=1}^{n} \varphi_j^*(z) \cdot \prod_{k \neq j} \Phi_k^*(z). \tag{4}$$

Proof. We may write $\delta^* = \max_{1\le j\le n}|r_j|$, where $r_j := \frac{y_j - X_j b}{|X|_j \cdot |b|}$; by convention, the value of a fraction with a zero denominator is zero. Observe that r_j's are independent normal variables with zero means and standard errors $\frac{\sigma}{|X|_j \cdot |b|}$. Hence the cumulative distribution function of $|r_j|$ is Φ_j^*. By independence of r_j's,

$$\Pr[\delta^* \le z] = \Pr[\max_j |r_j| \le z] = \Pr[(\forall j \in \{1,\dots,n\})|r_j| \le z] = \prod_{j=1}^{n} \Phi_j^*(z).$$

Differentiating this expression we obtain (4). □

Example. Consider the following data [12]:

j	1	2	3	4	5	6	7	8	9	10
x_j	2	4	6	8	10	12	14	16	2	16
y_j	14	16	14	18	18	22	18	22	4	32

Let us fit the model $E(y_j) = b_0 + b_1 x_j$ using two estimators. The first is OLS: we get $n^{[1]} = 10$, $\widehat{b}^{[1]} = (8.1233, 1.0752)^{\mathrm{T}}$ and $\widehat{\sigma}^{[1]} = 4.3603$. Now let us use another estimator: an OLS estimator equipped with a procedure for detection and exclusion of outliers. At the 'first sight', the outliers are $j = 9$ and $j = 10$. If we exclude them, we get the OLS estimate $n^{[2]} = 8$, $\widehat{b}^{[2]} = (12.9286, 0.5357)^{\mathrm{T}}$ and $\widehat{\sigma}^{[2]} = 1.793$.

Assume that estimated regression parameters are the true ones, i.e. $b = \widehat{b}^{[k]}$ for $k = 1$ and $k = 2$, respectively, and that $\sigma = \widehat{\sigma}^{[k]}$ for $k = 1$ and $k = 2$, respectively. The corresponding distributions φ_{δ^*} are plotted in Figure 1. Observe that in the case without outliers, the probability that $\delta^* \ge 1$ is negligible. This is interesting: note that if $\delta^* \ge 1$, then the tolerance interval for regression parameters contains zero which is an indicator that the regression model is not suitable for the data observed. The shapes of the distributions indicate that the model with outliers excluded fits the data significantly better. The Figure also shows that in the case without outliers, if $b = \widehat{b}^{[2]}$ and $\sigma = \widehat{\sigma}^{[2]}$, 'almost all' realizations of the random disturbances of the regression model lead to the quotient δ^* being below 40%.

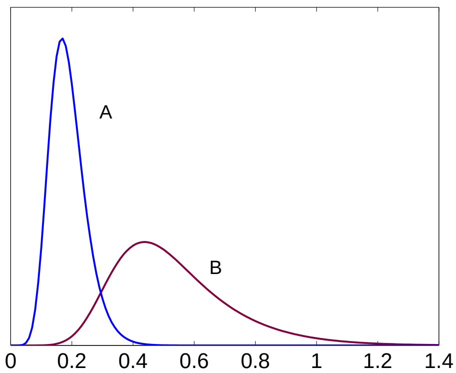

Figure 1: Distribution of δ^* for the model without outliers (A) and with outliers (B).

4 The confidence interval for δ^*

Intuitively, the OLS estimate $\widehat{b}$ is 'the best' estimate of the unknown value b. Having estimated $\widehat{b}$ by OLS, we compute δ^* as a function of $\widehat{b}$ as the exact value b is not known. Let us denote $\delta^*(\beta) = \max_{1\le j\le n} \frac{|y_j - X_j\beta|}{|X|_j \cdot |\beta|}$; now δ^* in (2) is $\delta^*(\widehat{b})$. Denote $\delta^*_{true} := \delta^*(b)$. Our aim is to construct an upper bound $\overline{\delta^*}$ for δ^*_{true} in terms of the *known* variables. Then, the interval $[0, \overline{\delta^*}]$ may be called a *confidence interval* for the unknown value δ^*_{true}.

Recall that the α-*confidence ellipsoid* for b, denoted $\mathcal{E}_\alpha$, is the least-volume ellipsoid centered at $\widehat{b}$ with the property that it covers the true value b with probability at least α. It is of the form $\mathcal{E}_\alpha = \{\beta : (\beta - \widehat{b})X^{\mathrm{T}}X(\beta - \widehat{b})^{\mathrm{T}} \le d \cdot \widehat{\sigma^2} \cdot F_{d,n-d}(\alpha)\}$, where $F_{d,n-d}$ stands for the F-distribution with d and $n - d$

degrees of freedom and $\widehat{\sigma^2}$ is the standard estimator of variance ($\widehat{\sigma^2} = \frac{1}{n-d}(y - X\widehat{b})^{\mathrm{T}}(y - X\widehat{b})$). It is interesting that under some sign-invariancy assumptions, it is possible to write down an easily-computable expression for $\overline{\delta^*}$.

Theorem 4. *Let $\alpha \in (0,1)$, let X be a positive matrix and let $\mathcal{E}_\alpha$ lie in the positive orthant of $\mathbb{R}^d$. Define*

$$\overline{\delta^*} := \max_{1\le j\le n} \max \left\{ \frac{y_j}{X_j\widehat{b} + (-1)^{1-\mathrm{sign}(y_j)} \cdot \sqrt{\frac{d\cdot\widehat{\sigma^2}\cdot F_{d,n-d}(\alpha)}{X_j X_j^T}} \cdot X_j(X^TX)^{-1/2}X_j^T} - 1, \right.$$
$$\left. 1 - \frac{y_j}{X_j\widehat{b} + (-1)^{\mathrm{sign}(y_j)} \cdot \sqrt{\frac{d\cdot\widehat{\sigma^2}\cdot F_{d,n-d}(\alpha)}{X_j X_j^T}} \cdot X_j(X^TX)^{-1/2}X_j^T} \right\}.$$

Then $\delta^*_{true} \le \overline{\delta^*}$ with probability at least α. □

The proof is omitted; it is available by the authors. The sign-invariancy assumptions may be further relaxed; however, then there is no compact expression for $\overline{\delta^*}$. But the situation is not bad: the computation of $\overline{\delta^*}$ may be reduced to a family of (computationally quite easy) convex optimization problems.

Probabilistic properties of δ^*, both in crisp and interval models, are subject of further research.

Acknowledgement. The first author's research was supported by the Internal Grant Agency of the University of Economics, Prague, 2010.

References

[1] Antoch, J., Hušková, M., and Jarušková, D.: *Off-line statistical process control.* In: *Multivariate Total Quality Control.* Physica-Verlag, Heidelberg, 2002, 1–86.

[2] Hesmaty, B., and Kandel, A.: *Fuzzy linear regression and its applications to forecasting in uncertain environment.* Fuzzy Sets and Systems 15 (1985), 159–191.

[3] Hladík, M., and Černý, M.: *Interval regression by tolerance analysis approach.* Technical report, KAM-DIMATIA Series 963 (2010).

[4] Hladík, M.: *Tolerance analysis in linear programming.* Technical report, KAM-DIMATIA Series 901 (2008).

[5] Hladík, M., and Černý, M.: *New approach to interval linear regression.* In: Kasımbeyli R. et al. (eds.): *24th Mini-EURO Conference On Continuous Optimization and Information-Based Technologies in The Financial Sector MEC EurOPT 2010. Selected Papers.* Technika, Vilnius, 2010, 167–171.

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

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

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

[9] Jun-Peng, G., and Wen-Hua., L.: *Regression analysis of interval data based on error theory.* In: *Proceedings of 2008 IEEE International Conference on Networking, Sensing and Control.* ICNSC, Sanya, China, 2008, 552–555.

[10] Rohn, J.: *A handbook of results on interval linear problems.* Czech Academy of Sciences, Prague, 2005. Available at: `http://uivtx.cs.cas.cz/~rohn/handbook/handbook.zip`.

[11] Tanaka, H.: *Fuzzy data analysis by possibilistic linear models.* Fuzzy Sets and Systems 24 (1987), 363–375.

[12] Tanaka, H., and Lee, H.: *Interval regression analysis by quadratic programming approach.* IEEE Transactions on Fuzzy Systems 6 (1998), 473–481.

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

[14] Sugivara, K., Ishii, H., and Tanaka, H.: *Interval priorities in AHP by interval regression analysis.* European Journal of Operational Research 158 (2004), 745–754.

[15] Ziegler, G.: *Lectures on polytopes.* Springer Verlag, 2004.