

Hansen-Bliek-Rohn Method Applied On Total Least Squares with Interval Data Input

¹Stephen Ehidihamhen Uwamusi

¹ Department of Mathematical Sciences, Kogi State University, Anyigba, Kogi State, Nigeria

-----ABSTRACT-----

In this paper, the method of Total least squares is presented for the set of statistical data points which are under interval uncertainty and, Hansen-Bliek-Rohn method, was applied to solve the resulting interval linear system with guaranteed inclusion bounds as demonstrated by numerical example. It is suggested that Hansen-Bliek-Rohn method always provide results which takes into account all round off errors which are as good as worst case error bounds with less computational efforts where theoretical floating point result fails.

Keywords: total least squares, Hansen-Bliek-Rohn method, circular interval arithmetic, worst case error bounds, floating point arithmetic MSC (2000): 65G20, 65G30, 65G40, 65 F20

Date of Submission: 23, August, 2013



Date of Acceptance: 30, October, 2013

I. INTRODUCTION

The paper presents interval methods on Least squares equation for interval data type problems. When experimental problems are expressed in terms of uncertainties, the need to provide bounds for the solution set can be achieved through the use of interval arithmetic. Practically, it is that in a situation where safety is of paramount importance, interval arithmetic is able to provide the worst case guarantee [1] where traditional floating point method fails. However interval arithmetic can be weak due to dependency problem which can be traced to a result of locality reasoning inheritable from arc consistency [2] and [3]. Interestingly this has been overcome in recent time [1].

Applications of Total least squares to problems can be found in signal processing, automatic control, system theory, various engineering practices, statistics, physics, economics, biology and medicine and a host of others.

Statistical approach has long been in use for describing how observed set of data points $(T_1, Y_1), (T_2, Y_2), \dots, (T_n, Y_n)$ of sample size n of a population fits the data but when the data points are expressed in the form of uncertainty any computation using traditional approach will undermine the quality of results. Thus interval arithmetic becomes a useful tool in this case. There exist five parameters $\mu_T, \mu_Y, \delta_T, \delta_Y$ and ρ for describing the behaviour of the statistical data. One of such ways [4] begins with elementary statistical analysis for traditional floating point approach such that the estimate for ρ could be defined by the use of product moment

$$\rho = E(T - \mu_T)(Y - \mu_Y) \quad (1.1)$$

The parameters μ_T, μ_Y can be estimated by \bar{T} and \bar{Y} where \bar{T} denotes $\frac{\sum_{i=1}^n T_i}{n}$ as the average of variables T_i . The

product moment is then written as $\frac{1}{n} \sum_{k=1}^n (T_k - \bar{T})(Y_k - \bar{Y})$.

Assuming δ_T and δ_Y are approximated by S_T and S_Y and denoting the sample estimate ρ as r , then the correlation coefficient is written in the form:

$$r = \frac{\sum_{k=1}^n (T_k - \bar{T})(Y_k - \bar{Y})}{n S_T S_Y} = \frac{n \sum TY - \sum T \sum Y}{\sqrt{(n \sum T^2 - (\sum T)^2)(n \sum Y^2 - (\sum Y)^2)}} \quad (1.2)$$

The z value for the normal distribution based on the knowledge of value of ρ is given in the form

$$z = \frac{1}{2} \log_e \frac{1+r}{1-r}, \quad (1.3)$$

and this gives the random variable z to be distributed with mean

$$\mu_z = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}, \text{ and standard deviation, } \delta_z = \frac{1}{\sqrt{n-3}}.$$

We move to discuss the computations of mean and variance for the statistical data when such data are subjected to interval uncertainty. As in [5], we describe method for computing mean and variance with interval data as follows:

First, we note that, $\mu = \frac{1}{n} \sum_i T_i$ is a monotonically increasing function of each n variables T_1, T_2, \dots, T_n and, the value for the interval mean was computed in the

$$\text{form: } \mu = \frac{1}{n} \left(T_1 + T_2 + \dots + T_n \right), \quad \bar{\mu} = \frac{1}{n} \left(\bar{T}_1 + \bar{T}_2 + \dots + \bar{T}_n \right).$$

Unfortunately, the same cannot be said of variance as it is dependent on T_i , its monotonicity fails in this circumstance as a result of dependency problem caused by locality reasoning inheritable from arc consistency. Besides, its computational complexity grows exponentially as the size n of data increases. It follows that the true

estimate of the statistic $V(T_1, T_2, \dots, T_n)$ on the interval $\{ [T_1, \bar{T}_1], \dots, [T_n, \bar{T}_n] \}$, will be best described as

below. First recalling that midpoint for interval $T_i = [T_i, \bar{T}_i]$ is written as $T_{ic} = \frac{1}{2} (\bar{T}_i + T_i)$, and its

radius $\Delta_i = (\bar{T}_i - T_i) / 2$. Following [6], the range of V could be written as $[V_0 - \Delta, V_0 + \Delta]$, where

$$\Delta = \sum_{i=1}^n |V_i| \Delta_i, \quad V_i = \frac{\partial V}{\partial t_i} = \frac{2}{n} \left(T_{ic} - \bar{\mu} \right), \text{ and } \bar{\mu} \text{ is the average midpoints of}$$

$$T_i, \quad V_0 = \frac{1}{n} \sum_{i=1}^n \left(T_{ic} - \bar{\mu} \right)^2 \text{ is the average variance of the midpoint values } T_{1c}, \dots, T_{nc}.$$

The radius of the variance Δ was defined as $\Delta = \frac{2}{n} \sum_{i=1}^n \left(\left| T_{ic} - \bar{\mu} \right| \right) \cdot \Delta_i$. The quadratic term for the variance

which was negligibly small in the truncated Taylor series was given as $\frac{1}{n} \sum_{i=1}^n ((\Delta T_i)^2 - (\Delta \mu)^2)$, where,

$$0 \leq \Delta \mu = \frac{1}{n} \sum_{i=1}^n \Delta T_i. \text{ Thus the inclusion interval for variance V was computed}$$

$$\text{as } V = [V_0 - \Delta, V_0 + \Delta + \Delta^{(2)}].$$

The remaining section in this paper is structured as follows. Section 2 gives the introduction of the problem of Least squares as it relates Co-variance matrix and the distribution of variance as a biased estimator. In section 3, the extent to which an interval matrix that is not strongly regular can become singular if an attempt is made to extend the radius of an interval matrix by a certain factor based on the idea given in [7] was the main point of focus. Section 4 gives the numerical method we used to solve the resulting interval linear system obtained from the total least squares when the variable in the equation were subjected to uncertainty either due to contamination of the measuring instruments used in the model or some noise coming from the inexact observation of the model. We used the Hansen-Blik-Rohn method which can be found in [8,9 and 10] to solve the Linear interval system of equations. Section 5 gives numerical example of the presented methods. Section 6 gives conclusion of the paper based on our findings.

II. GENERAL LINEAR CASE

Consider the general linear model

$$y_i = \beta_1 T_{i1} + \beta_2 T_{i2} + \dots + \beta_p T_{ip} \quad 2.1$$

Where T_{ij} is the j-th observation on the j-th independent variable and the first variable T_{i1} takes value 1. The residual is given by

$$e_i = Y_i - (\hat{\beta}_1 T_{i1} + \hat{\beta}_2 T_{i2} + \dots + \hat{\beta}_p T_{ip}) \quad (2.2)$$

The normal equation is obtained in a general way by minimizing the sum of least squares error:

$$\sum_{i=1}^n \sum_{k=1}^p T_{ij} T_{ik} \hat{\beta}_k = \sum_{i=1}^n T_{ij} Y_i, \quad j = 1, 2, \dots, p \quad (2.3)$$

Where T is $m \times n$ matrix, Y is an $m \times 1$ vector, β is $n \times 1$ vector.

In matrix form, we have a representation

$$(T' T) \hat{\beta} = T' Y \quad (2.4)$$

The condition number for the rectangular matrix T with full column rank is discussed in terms of 2-norm and is in the sense of [11], given by

$$T \in R^{m \times n}, \text{rank}(T) = n \Rightarrow k_2(T) = \frac{\delta_{\max}(T)}{\delta_{\min}(T)}, \text{ where } \delta(T) \text{ denotes the singular value of } T.$$

Considering that matrix $T' T \in A$ for real matrix $A \in R^{n \times n}$ we discuss the statistical interpretation of equation 2.4 as follows:

$$E(\hat{\beta}) = E\left\{(T' T)^{-1} T' y\right\} = (T' T^{-1}) T' T \beta = \beta. \quad (2.5)$$

The sampling variance $\text{var}(\hat{y}) = \delta^2 I_n$

$$\text{It can be showed that } \hat{\delta}^2 \text{ is a biased estimator of } \delta^2, \text{ and } E\left(\hat{\delta}^2\right) = \frac{n-m}{n} \delta^2. \quad (2.6)$$

Definition 2.1: Assuming that $\hat{\beta} = A^{-1} T'$,

$$\hat{\delta}^2 = \frac{1}{n-m} \left(Y - T \hat{\beta} \right)' \left(Y - T \hat{\beta} \right) = \frac{1}{n-m} \left(Y' (I - T A^{-1} T') Y \right), \text{ denoting } G = (I - T A^{-1} T'), \text{ the}$$

parameters $\hat{\beta}$ and $\hat{\delta}^2$ are said to be independent if $(A^{-1} T') G = O$.

As is well known, distribution of $\hat{\delta}^2$ is discussed as follows. Since, $\hat{\delta}^2 = \frac{\left(Y - T \hat{\beta} \right)' \left(Y - T \hat{\beta} \right)}{n-m}$, then it

follows that $(n-m) \hat{\delta}^2 = \left(Y - T \hat{\beta} \right)' \left(Y - T \hat{\beta} \right) = Y' (I - T A^{-1} T') Y$. But $(I - T A^{-1} T')$ is an

idempotent matrix as a result, it holds that $(n-m) \hat{\delta}^2 = Y' (I - T A^{-1} T') Y = Y' G Y$ which gives

$$\hat{\delta}^2 = \frac{Y' G Y}{n-m}.$$

That is, $\frac{(n-m) \hat{\delta}^2}{\delta^2} \sim \chi_{n-m, \lambda}^2$ with non- centrality parameter $\lambda = \frac{\beta' T' G T \beta}{2 \delta^2}$. As $\lambda \rightarrow 0$, this gives Chi-square distribution.

III. COMPONENT WISE DISTANCE AS A CONDITION NUMBER TO THE WEIGHT MATRIX E

Considering the fact that we are interested in a situation when the coefficients of $T' T \in A$ are interval entries where, $A \in IR^{n \times n}$ and a non negative $E \in R^{n \times n}$. It should be noted that for interval matrix

$A = [A_c - E, A_c + E]$ it is that:

$$A_c = \frac{1}{2} \begin{pmatrix} A^+ & A^- \\ A^- & A^+ \end{pmatrix}, \quad E = \frac{1}{2} \begin{pmatrix} A^- & A^- \\ A^- & A^- \end{pmatrix}.$$

Given knowledge of $\delta(A, E)$ as being defined as a number $\alpha = \delta(A, E)$ for which the matrix

$[A - \alpha \cdot E, A + \alpha \cdot E]$ is singular, and setting $\omega = \rho(|A^{-1}| \cdot E)^{-1}$ we define any matrix as

$[A - \omega' \cdot E, A + \omega' \cdot E]$, (where $\omega' < \omega$) is strongly regular. In [12] it was proved that the ratio

$\frac{\delta(A, E)}{\rho(|A^{-1}| \cdot E)^{-1}}$ is the distance between strong regularity and regularity of the matrix A. Thus, Bauer-Skeel

condition number written as $Cond_{bs}(A, E) = \left\| |A^{-1}| \cdot E \right\|$ for some norm $\| \cdot \|$ relates component wise – distance as a condition number to the weight matrix E. A bound where such a matrix A could become singular

occurs, assuming $\frac{1}{\rho(|A^{-1}| \cdot E)} < \delta(A, E)$, when:

$$\frac{1}{\rho(|A^{-1}| \cdot E)} \leq \delta(A, E) \leq (3 + 2\sqrt{2}) \cdot n \cdot \frac{1}{\rho(|A^{-1}| \cdot E)}. \quad (3.1)$$

It follows that equation 3.1 states that, there exists a bound $n \leq \eta(n) \leq (3 + 2\sqrt{2}) \cdot n$ for which an interval matrix that is not strongly regular will produce a singular matrix if an attempt is made to increase radius by at most $(3 + 2\sqrt{2}) \cdot n$. As advised in [7], the factor

$$(3 + 2\sqrt{2}) \cdot n \cdot \frac{1}{\rho(|A^{-1}| \cdot E)}, \text{ cannot be replaced by the term } (3 + 2\sqrt{2}) \cdot \frac{1}{\rho(|A^{-1}| \cdot E)}.$$

Assuming equation 2.4 is in normal form which can be written in the linear interval system $\mathbf{A} \beta = \mathbf{b} = \{ \beta : A \in \mathbf{A}, b \in \mathbf{b} \}$, it is assumed that $T' T \in \mathbf{A}$, and $T' Y \in \mathbf{b}$ then we aim to solve for the unknown interval $\beta \in \mathbb{R}^n$ in order to obtain the Hull of the solution set

$$\sum (\mathbf{A}, \mathbf{b}) = \left(\hat{\beta} \in \mathbb{R}^n \mid A \in \mathbf{A} \exists b \in \mathbf{b} \ni \mathbf{A} \hat{\beta} = \mathbf{b} \right). \quad (3.2)$$

Several characterizations of solution sets to (3.2) exist, for example, [13], gave three types of such solution sets to include among others as follows:
a tolerable solution set

$$\sum_{\forall \exists} (\mathbf{A}, \mathbf{b}) = \left\{ \hat{\beta} \in \mathbb{R}^n \mid (\forall A \in \mathbf{A})(\exists b \in \mathbf{b}) \left(\mathbf{A} \hat{\beta} = b \right) \right\}, \quad (3.3)$$

The controllable solution set

$$\sum_{\forall \exists} (\mathbf{A}, \mathbf{b}) = \left\{ \hat{\beta} \in \mathbb{R}^n \mid (\forall b \in \mathbf{b})(\exists A \in \mathbf{A}) \left(\mathbf{A} \hat{\beta} = b \right) \right\}, \quad (3.4)$$

and the united solution set

$$\sum_{\exists \exists} (\mathbf{A}, \mathbf{b}) = \sum_{\exists \exists} \left\{ \hat{\beta} \in \mathbb{R}^n \mid (\exists A \in \mathbf{A})(\exists b \in \mathbf{b}) \left(\mathbf{A} \hat{\beta} = b \right) \right\}. \quad (3.5)$$

The terms \forall, \exists appearing above are all quantifiers.

IV. THE INCLUSION METHODS FOR EQUATION 2.4

The inclusion method under consideration for the solution of equation 2.4 will be the adoption of Hansen-Blik-Rohn [10] and, where united solution set is used in our consideration.

The enclosure for the inverse interval matrix $A = [A_c - E, A_c + E]$, provided that $\rho(|A_c^{-1}| \cdot E) < 1$ in the sense of [9 and 10] may be written as:

$$A^{-1} \subseteq [\min \{ \underline{B}, T_v \underline{B} \}, \max \{ \bar{B}, T_v \bar{B} \}] \quad (4.1)$$

Where

$$M = (I - |A_c^{-1}|E)^{-1},$$

$$\mu = (M_{ii})^T, i = 1, 2, \dots, n,$$

$$T_v = (2T_\mu - I)^{-1},$$

$$\underline{B} = -M |A_c^{-1}| + T_\mu (A_c^{-1} + |A_c^{-1}|),$$

$$\bar{B} = M |A_c^{-1}| + T_\mu (A_c^{-1} - |A_c^{-1}|)$$

In equation 4.1, T_v is an orthant which is very crucial for the existence of Hansen-Blik-Rohn method, see e.g., [8] for an overview, provided $\rho(|A^{-1}|.E) < 1$.

The error bound for the computed approximate solution β^* to the true solution β provided by using Rohn's method 4.1, as described in [10] is distributed as follows.

$$\beta - \beta^* = (T'T)^{-1} T'T\beta - (T'T)^{-1} T'Y$$

$$= (T'T)^{-1} (T'T\beta - T'Y) \tag{4.2}$$

By taking norm of both sides of equation 3.2 in the sense of [14], gives that

$$\|\beta - \beta^*\|_2 \leq \|(T'T)^{-1}\|_2 \cdot \|T'(T\beta - Y)\|_2 \leq \alpha \|T'(T\beta - Y)\|_2 \tag{4.3}$$

Using right hand side of equation 4.2 we have

$$\|(T'T)^{-1}\|_2 \cdot \|T'(T\beta - Y)\|_2 = \|(T'T)^{-1}\|_2 \cdot \|T'(T\beta - T\beta^* + T\beta^* - Y)\|_2$$

$$= \|(T'T)^{-1}\|_2 \cdot \|T'T(\beta - \beta^*)\|_2$$

$$\approx K(T)^2 \cdot \|\beta - \beta^*\|_2$$

The error bound $\|\beta - \beta^*\|$ is directly proportional to $K(T)$, this is more so, since, $\alpha \|T'(T\beta - Y)\|_2$ is the power 3 of the condition number $K(T)$ of the matrix.

V. IMPLEMENTATION AND EXPERIMENTATION

Problem 1.

Consider the given problem taking from [15] of regressing y on t

" $Y = \beta_0 \ln T + \beta_1 \cos T + \beta_2 e^T$ ". The equation is

$$\varphi(\beta_0, \beta_1, \beta_2) = \sum_{j=1}^m (\beta_0 \varphi(\ln T_j) + \beta_1 \varphi(\cos T_j) + \beta_2 \varphi(e^{T_j}) - Y_j)^2 \tag{5.1}$$

The data points are

T	Y
0.24	0.23
0.65	-0.26
0.95	-1.10
1.24	-0.45
1.73	0.27
2.01	0.10
2.23	-0.29
2.52	0.24
2.77	0.56
2.99	1.00

Take data noise to be $\varepsilon = 1\%$ for each data set (T_j, Y_j) . Result obtained for normal equations without noise on the data using MATLAB 2007 Version is $(-1.04103, -1.26132, 0.03073)^T$

Result obtained using Hansen-Blik- Rohn’s method (3.1) for $(\beta_0, \beta_1, \beta_2)^T$ where the data under investigation were subjected to some noise of $\varepsilon = 1\%$ was found to be

$$\left([-1.0312, -1.0085], [-1.2757, -1.1951], [0.0305, 0.0309] \right)^T.$$

Therefore the regression equation is

$$Y(T) = [-1.0312, -1.0085] \varphi(\ln T) + [-1.2757, -1.1951] \varphi(\cos T) + [0.0305, 0.0309] \varphi(e^T),$$

Let us note that the solution set given by using Hansen-Blik-Rohn’s method 3.1 is closed in the sense of Oettli-Prager theorem [13]. It should also be noted that the mapping $(A, b) \rightarrow (A^T A)^{-1} A^T b$ is also compact and continuous in the Least squares sense.

By taking the value of t to be in the interval of [3.0, 3.5] we have the following result in midpoint-radius interval, the value for equation 4.2 in midpoint-radius interval was computed to be

$$Y[T] = [-1.508421816, 0.444770179].$$

In the same reasoning, the value of $\varphi(\beta_0, \beta_1, \beta_2) = -1.693769447$ for the midpoint interval [3.0, 3.5] was computed.

VI. CONCLUSION

The paper considered Total Least squares method for statistical data set which is under interval uncertainty. Hansen-Blik-Rohn’s method was applied on the Least squares method which provides tight inclusion bounds for the Least squares problem with less computational efforts. It was demonstrated in our example that computed result using Hansen-Blik-Rohn method showed a close relationship with guaranteed worst case error bound which takes into account all round off errors as compared with theoretical floating point result. We also examined regularity condition of resulting interval matrix coming from system of equation 2.4 taking into consideration distance between strong regularity and regularity of the interval matrix A in a sense analogous to [7] where Bauer-Skeel condition number played a very crucial role which relates component wise –distance as a condition number to the weight matrix E .

REFERENCES

- [1] Neumaier and Pownuk(2006). Linear systems with large uncertainties with applications to truss structures. *Reliable Computing* Vol.13 issue 2,149-172.
- [2] Uwamusi S.E (2011). Effects of repeatedly used preconditioner on computational accuracy for nonlinear interval system of equation. *International Journal of the Physical Sciences* Vol. 6(33), pp. 7505-7511. <http://academicjournals.org/IJPS>
- [3] Kearfott RB and Xing Z (1993). An interval step control for continuation methods, *SIAM J. Numer. Anal.* 0(0):1-23.
- [4] Hoel P (1971). *Introduction to Mathematical statistics* fourth edition, John Willey and Sons, New York.
- [5] Kreinovich V and Nguyen HT (2007). Online algorithms for computing mean and variance of interval data and their use in intelligent systems, *Information Science*, 177(16), 3228-3238.
- [6] Xiang G (2007). Fast algorithms for computing Statistics under interval uncertainty, with applications to Computer Science and to Electrical and Computer Engineering, Ph.D Thesis, Department of Computer Science, The University of Texas at El Paso.
- [7] Rump SM (1999). Ill-conditioned matrices are component wise near to singularity, *SIAM Review* 41(10), 102-112.
- [8] Rohn J (2005). *A Handbook of results on interval linear problems*, Czech Academy of Sciences, Prague, Czech Republic, European Union, www.cs.cas.cz/rohn
- [9] Rohn J., (2010a). A Characterization of Strong regularity of interval Matrices. Technical report No V. 1072. Institute of Computer Science , Academy of Sciences of the Czech Republic, PP. 1-7.
- [10] Rohn J., (2010b). An improvement of the Bauer-Skeel Bounds. Technical Report No. V- 1065. Institute of Computer Science Academy of Sciences of Czech Republic, PP. 1-9.
- [11] Golub G and Van- Loan F (1983). *Matrix Computations*. The Johns Hopkins University Press, Baltimore Maryland.
- [12] Rump SM (2003). Optimal scaling for p-norms and componentwise distance to singularity. *IMA Journal of Numerical Analysis (IMAJNA)*, Vol. 23, 1-9.
- [13] Shary SP (2002). A new technique in systems analysis under interval uncertainty and ambiguity. *Reliable Computing* 8:321-418.
- [14] Miyajima S (2010). Fast enclosure for solutions in underdetermined systems. *Journal of Computational and Applied Mathematics (JCAM)*, 234:3436-3444.
- [15] Cheney W and Kincaid D (1980). *Numerical Mathematics and Computing*, Brooks/ Cole Publishing Company Monterey, California.