Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 123-132
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.4917

Goals programming multiple linear regression model for optimal estimation of Electrical Engineering staff according to load Demand

Shrook A. S. Al-Sabbah [a], Zainab Hassan Radhy [b*], Hayder Al Ibraheemi[c]

[a] 1Department of Statistics, Administration and Economic College, Karbala University, Iraq.
[b] 2College of computer science and information technology - University of Al-Qadisiyah - Iraq.
[c] 3Faculty of Technical Engineering/Islamic University-Iraq.

(Communicated by Dr. Ehsan Kozegar)

Abstract

The linear multiple regression model is one of the prediction models whose parametric estimations could be achieved in different methods. Ordinary Least Square error (OLS) is most popular in this field of application. Although it could accurately achieve the estimation task, it fails in processing the multiple objective models. From the other side of view, the load demand for electrical energy continuously rises around the world. The governments always tackle the increase in electrical load demanding by establishing more electrical power plants and more power distribution directories. Future prediction for several electrical engineers to manage and provide technical supports for these plants and directories becomes, nowadays, urgent. This paper addresses the estimating drawback of OLS by employing the Goals Programming (GP) in the field of parametric estimation. The validity of the proposed method was applied to estimating the required number of electrical engineers, in the next coming years, as the electrical load considerably increases. Thereby, the GP was used, in this work, to determine the best linear representation for a set of data. The obtained results proved that the (GP) method is more flexible and efficient in dealing with such subject area especially in the case of multiple objective models.

Keywords: Linear multiple module, Goals programming method, least-square error, Electrical load demand, Electrical engineers number

*Corresponding Author: Zainab Hassan Radhy

Email address: Shorouk.a@uokerbala, Zainb.hassan@qu.edu.iq, hayder.mazin@iunajaf.edu.iq (Shrook A. S. Al-Sabbah [a], Zainab Hassan Radhy [b*], Hayder Al Ibraheemi[c])

Received: October 2020 *accepted:* February 2021

1. Introduction

For any statistician, it is important to study the regression analysis and analyzing the effect of an independent variable on the dependent variable. The goals programming (GP) involves solving the problem of one specific objective function. Moreover, FP is also applicable in manipulating a set of goals. In linear and integer programming methods, the objective function is measured only in one dimension. In contrast, by goals programming, conflicting goals or goals with different priorities and weights could be combined and solved by using the simplex algorithm. In this paper, we will focus on the multiple regression linear analysis and the methods of estimating the module parameters. One of these methods is the classical method, least-square error (OLS) [5], [9].

Reference [2] used linear programming, LP, to establish models for solving real-time industrial problems. The trend of this reference was in two directions, firstly by using the models of LP as a guide for data collecting and secondly by applying the LP models in analyzing the fruitful areas. Reference [3] presented a survey of recent developments in goal programming and multiple objective optimizations. Special attention was paid toward goal programming, particularly, goal interval programming along with an example that applied sequences of ordinary linear programming problems in solving certain nonlinear problems. A planning model for building a workforce was presented by reference [4]. The model was applied for data from an electrical distribution establishment. It was mentioned by reference [1] although the least-squares method is the most frequently used procedure for estimating the regression model parameters, it appears nonbiased outcomes in presence of outliers. Accordingly, the reference presented a goal programming approach to estimate parameters of a regression model when the analysis includes outliers. Reference [8] introduced a comparative study among three linear programming models each to estimate the regression parameters utilizing the method of least absolute values. It was found that the three model performances differ concerning several observations.

The proposed method by this article proves the possibility of utilizing the goals programming (GP) in the parametric estimation of the multiple linear regression model. The paper is organized into eight sections. After the abstract, the introduction explores the main paper topic and a brief review of the relevant works. Section two presents the aim of the paper study. A theoretical presentation for the multiple linear models is given in section three. Section four introduces the proposed goal programming model, while the parametric estimation by multiple linear regression is detailed in section five. To show the validity of the proposed work, real-time data are applied to the proposed GP model in section six. The results and conclusions are discussed in sections seven and eight respectively.

2. Aim of the study :

This article applies a linear programming approach to estimate a parametric linear multiple regression model by using the goals programming method. It is known that the GP method is limited to two variables (three parameters), but in this paper, it was applied for seven variables (eight parameters). Moreover, the least square error method was also applied in this work to compare the two approaches and to prove the validity of the proposed method.

3. Multiply regression Linear model:

In this model, suppose that the variable y expresses the dependent variable, and the variables $(x_1, x_2, \ldots . x_k)$ refer to (k) of independent variables which draw from (n) observations number. Then the following linear function can be obtained:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_k x_{ik} + \epsilon_i + \cdots\cdots \quad (1)$$

where $i = 1, 2, \ldots n$ And ϵ_i represents random errors that distribute normally with zero mean and σ^2 variance.

To simplify the mathematical operations, it is convenient to re-express the linear function, given in (1), in following matrix form :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \cdots\cdots \quad (2)$$

Or shortly by:

$$Y = XB + U \cdots\cdots (2) \cdots$$

where:

$Y : (n * 1)$ is the dependent variables column vector.

$X : (n * (k + 1))$ is the matrix of independent variables.

B : is the regression parameters column vector $((k + 1) * 1)$

U : is the error column vector $(n * 1)$

Some assumed effects on the linear multiple regression module are listed below:

- The observation values are represented by the elements of the matrix (X).
- No linear relationship between the independent Variables, i.e. Ranks $(X) = (k + 1) < n$.
- The columns matrix is linear independent to error random vector (U), $cov(X, U) = E(X'U) - [E(X)]'[E(U)] = 0$.
- Random Error $(\epsilon_i, i = 1, 2, \ldots, n)$ has normal distribution $\epsilon_i \sim N(0, \sigma^2)$

4. Goals Programming Module:

The goals programming (GP) technique has become a widely used approach in the field of operations research. GP was firstly developed and introduced by Charnes and Cooper [2], [3] in 1960. The development of this solution approach was extended by Ijiri [6]. The goal programming method is an improved method for solving multiple objective problems. The main idea of this approach is to determine the priority and weight of each goal (P_k). Then, it searches to find a solution that minimizes the total weighted function of deviations. In this method, two types of variables, decision variable and deviation variable, could be expressed. Therefore, the goals programming is given as follow:

$$Min\ a = \{P_1(d_1^+, d_1^-), P_2(d_i^+, d_i^-), \ldots\ldots P_k(d_i^+, d_i^-)\} \cdots \quad (4)$$

Subject to:

$$\sum_{j=1}^{k} a_{ij}x_j = b_i + d_i^+, d_i^- fori = 1, 2, \dots, n$$
$$x_j, d_i^+, d_i^- \geqslant 0$$

Where:

a : Goal function
P_k : Priority number i in goal function
x_j : Decision Variable
a_{ij} : Factor decision Variable to number j in goal i.
d_i^+ : Positive déviation variable.
d_i^- : Negative deviation variable.
b_i : The value of the goal function.

5. Parametric estimation by multiple linear regression:

The key point behind the regression analysis is to utilize a known past data to create an estimation function that mathematically links a dependent and independent variable(s). Then, the estimation equation is then employed to predict the dependent variable future values [1]. In this paper, one of the most and commonly used classical methods, the ordinary least square (OLS), is applied to find the parametric estimation. It is based on the process of finding a straight line which goes through a set of observation in a certain way so that the sum of the squares of differences between the observation points and the line is minimum. This paper used this method to estimate the multiple parametric coefficients $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ of equation (2) and re-write as the following [7]:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \cdots\cdots \quad (5)$$
$$Y = XB + U \cdots\cdots (6)$$
$$U = Y - XB \cdots\cdots (7)$$

By squaring both sides of eq. (7), yields:

$$Q = U'U = (Y_X B)'(Y - XB) \cdots\cdots (8)$$

Expanding eq. (8) gives:

$$Q = U'U = Y'Y - 2B'X'Y + B'X'XB \cdots\cdots (9)$$

To find the value of B which minimizes $U'U$, the derivative ratio to B_j is taken and then set to zero.

$$\frac{\partial Q}{\partial B} = \begin{bmatrix} \frac{\partial Q}{\partial B_0} \\ \frac{\partial Q}{\partial B_1} \\ \frac{\partial Q}{\partial B2} \\ \cdots \\ \frac{\partial Q}{\partial B_k} \end{bmatrix} = -2X'Y + 2X'X\hat{B} \cdots\cdots (10)$$

Eq. (10) can be written as the following:

$$X'X\hat{B} = X'Y\hat{B} = (X'X)^{-1}X'Y \cdots\cdots (11)$$

] where $\hat{B}$ represents the estimation vector by the ordinary least square. The result of eq. (10) can be rewritten more clearly as the following:

$$\frac{\partial Q}{\partial B_0} = (-2)\sum_{i=1}^{n} y_i - (\hat{B}_0 + \hat{B}_1 x_{1i} + \hat{B}_2 x_{2i} + \cdots + \hat{B}_k x_{ki}) = 0$$
$$\frac{\partial Q}{\partial B_1} = (-2)\sum_{i=1}^{n} x_{1i}[y_i - (\hat{B}_0 + \hat{B}_1 x_{1i} + \hat{B}_2 x_{2i} + \cdots + \hat{B}_k x_{ki}) = 0]$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\frac{\partial Q}{\partial B_k} = (-2)\sum_{i=1}^{n} x_{ki}[y_i - (\hat{B}_0 + \hat{B}_1 x_{ki} + \hat{B}_2 x_{ki} + \cdots + \hat{B}_k x_{ki}) = 0] \cdots\cdots (12)$$

By solving equation (12):

$$\sum_{i=1}^{n} y_i = \hat{B}_0 \sum_{i=1}^{n} 1 + \hat{B}_1 \sum_{i=1}^{n} x_{1i} + \hat{B}_2 \sum_{i=1}^{n} x_{2i} + \cdots + \hat{B}_k \sum_{i=1}^{n} x_{ki}$$
$$\sum_{i=1}^{n} x_{2i} y_i = \hat{B}_0 \sum_{i=1}^{n} x_{2i} + \hat{B}_1 \sum_{i=1}^{n} x_{1i} x_{2i} + \hat{B}_2 \sum_{i=1}^{n} x_{2i}^2 + \cdots + \hat{B}_k \sum_{i=1}^{n} x_{2i} x_{ki}$$
$$\sum_{i=1}^{n} x_{ki} y_i = \hat{B}_0 \sum_{i=1}^{n} x_{ki} + \hat{B}_1 \sum_{i=1}^{n} x_{1i} x_{ki} + \hat{B}_2 \sum_{i=1}^{n} x_{2i} x_{ki} + \cdots + \hat{B}_k \sum_{i=1}^{n} x_{ki}^2 \cdots\cdots (13)$$

Eq. (13) could be used in the goals programming approach to estimate the coefficients of multiply linear parametric regression $(\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ as following:

Suppose:

$$\hat{B_{i-1}} = x_i, y_i = b, \epsilon_i = d_i^+, d_i^-$$

Where:

ϵ_i : Represent the error in multiply linear regression model
d_i^+, d_i^- : Negative and positive deviations in model
y_i : Dependent variables in a multiple linear regression model
$\hat{B_{i-1}}$: Estimations in multiply linear regression model
x_i : Deviation variables in a multiple linear regression model

6. Application :

To implement the thoughts that have been mentioned in the theoretical part, a Matlab m-file was built to apply two different methods on the data given by table 1 [4] from a directory of Electricity Distribution. The aim is to predict the required number of employees of engineering staff to satisfy the directory policy. The number of engineering staff could be affected by several factors:

x_1 : The received electrical energy.
x_2 The sold electrical energy.
x_3 : Number of energy Consumers.
x_4 : Number of the energy power stations $33/11kv$
x_5 : Number of the $33kv$feeders.
x_6 : Number of the $11kv$feeders.
x_7 : The load demands

Several engineering staff in the directory of electricity distribution within the period (1985-1994) is shown in the table1 below:

Table 1: number of engineering between 1985 and 1994 [4]

years	Y : number of staff	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1985	101	5.122	4.627	542	86	167	869	1060
1986	116	5.659	4.349	580.6	90	179	925	1190
1987	114	6.330	5.255	609.65	92	184	939	1252
1988	119	6.545	5.595	638.7	95	191	959	1285
1989	133	7.312	6.216	668.3	96	202	998	1277
1990	134	8.046	7.001	678.64	97	206	1008	1354
1991	142	7.911	8.934	706.5	98	208	1046	1435
1992	143	5.114	4.490	720.5	99	208	1050	1312
1993	145	8.050	7.245	741.2	100	211	1077	1453
1994	128	8.768	7.253	759.2	101	211	1088	1534

The predictor model is given as follows:

$$y_i = \hat{B}_0 + \hat{B}_1 x_{i1} x_{2i} + \hat{B}_2 x_{i2} + \hat{B}_3 x_{i3} + \hat{B}_4 x_{i4} + \hat{B}_5 x_{i5} + \hat{B}_6 x_{i6} + \hat{B}_7 x_{i7} \cdots\cdots (14)$$

By using the data of table (1) and equation (14), the estimation model by (OLS) method is as shown in table 2.

Table 2: Parameters of the estimation model by (OLS) method

Parameter	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Value	-479.763	-13.093	13.9993	-0.92475	4.05582	0.66876	0.7692	-0.0419

Thereby, the predictor equation for number of engineers by (OLS) method could be written as follow:

$$\begin{aligned}\hat{y}_i =& - 479.763 - 13.093 x_{i1} + 13.9993 x_{i2} - 0.92475 x_{i3} + 4.05582 x_{i4} + 0.66876 x_{i5} + 0.7692 x_{i6} \\ &- 0.0419 x_{i7} \cdots\cdots (15)\end{aligned}$$

Re-estimating model parameters by goals programming utilizing equation (13) and data of table (1), then the goal programming linear form is given as the follow:

$$Min\ a = P_i \sum_{i=1}^{n} (d_i^+, d_i^-)$$

Subject to:

$$\begin{aligned}
&x_1 + 5.122x_2 + 4.627x_3 + 542x_4 + 86x_5 + 167x_6 + 869x_7 + 1060x_8 + d1 - d_1^+ = 101\\
&x_1 + 5.659x_2 + 4.349x_3 + 580.6x_4 + 90x_5 + 179x_6 + 925x_7 + 1190x_8 + d2 - d_2^+ = 116\\
&x_1 + 6.330x_2 + 5.255x_3 + 609.65x_4 + 92x_5 + 184x_6 + 939x_7 + 1252x_8 + d3 - d_3^+ = 114\\
&x_1 + 6.545x_2 + 5.595x_3 + 638.7x_4 + 95x_5 + 191x_6 + 959x_7 + 1285x_8 + d4 - d_4^+ = 119\\
&x_1 + 7.312x_2 + 6.216x_3 + 668.3x_4 + 96x_5 + 202x_6 + 998x_7 + 1277x_8 + d5 - d_5^+ = 133\\
&x_1 + 8.046x_2 + 7.001x_3 + 678.64x_4 + 97x_5 + 206x_6 + 1008x_7 + 1354x_8 + d6 - d_6^+ = 133\\
&x_1 + 7.911x_2 + 8.934x_3 + 706.5x_4 + 98x_5 + 208x_6 + 1046x_7 + 1435x_8 + d7 - d_7^+ = 142\\
&x_1 + 5.114x_2 + 4.490x_3 + 720.5x_4 + 99x_5 + 208x_6 + 1050x_7 + 1312x_8 + d8 - d_8^+ = 143\\
&x_1 + 8.050x_2 + 7.245x_3 + 741.2x_4 + 100x_5 + 211x_6 + 1077x_7 + 1453x_8 + d9 - d_9^+ = 145\\
&x_1 + 8.768x_2 + 7.253x_3 + 759.2x_4 + 101x_5 + 211x_6 + 1088x_7 + 1534x_8 + d10 - d_{10}^+ = 128\\
&\qquad x_j, d_i^+, d_i^- \geq 0\ where\ i = 1, 2, \cdots, 10, j = 1, 2, \cdots, 8
\end{aligned}$$

Solving the above model by the simplex method, the obtained results are:

Table 3: parameters of estimation model by goals programming method

Parameter	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Value	-457.459	-12.93974	13.91550	-0.90319	3.6223	0.8670	0.7317	-0.0399

Thereby, predictor equation by GP method is as the follow:

$$\begin{aligned}\hat{y}_i = &-457.459 - 12.93974x_{1i} + 13.91550x_{2i} - 0.90319x_{3i} + 3.6223x_{4i} + 0.8670x_{5i} + 0.7317x_{6i}\\ &- 0.0399x_{7i} \cdots\cdots (16)\end{aligned}$$

7. Results

Figure 1 shows how the value of the target varies with the progress of estimation iterations.

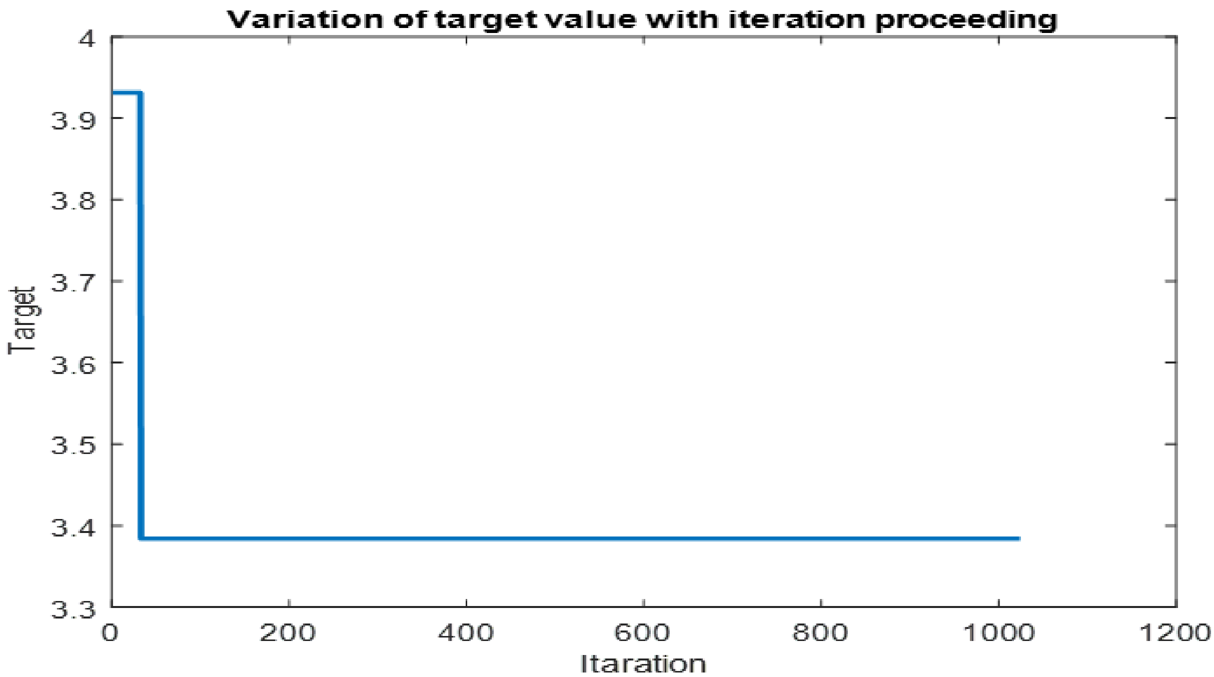

Figure 1: Variation of target value with iteration progressing

A comparison between the results of both methods, OLS and GP, could be achieved by using the analysis of variance (ANOVA). The following are the hypotheses of ANOVA:

$H_0 : \mu_1 = \mu_2$

H_1 :mean are not all equal, the level of significance is 0.05

Descriptives

degree

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
ols		8	-59.291262	170.0565650	60.1240751	-201.462109	82.879584	-479.7630	13.9990	
goals		8	-56.528000	162.1809410	57.3396216	-192.114660	79.058660	-457.5000	13.9155	
Total		16	-57.909631	160.5373877	40.1343469	-143.453967	27.634704	-479.7630	13.9990	
Model	Fixed Effects			166.1654190	41.5413547	-147.006976	31.187713			
	Random Effects				41.5413547[a]	-585.742590[a]	469.923327[a]			-3447.550497

a. Warning: Between-component variance is negative. It was replaced by 0.0 in computing this random effects measure.

Test of Homogeneity of Variances

degree

Levene Statistic	df1	df2	Sig.
.006	1	14	.939

ANOVA

degree

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.542	1	30.542	.001	.974
Within Groups	386553.250	14	27610.946		
Total	386583.793	15			

Figure 2: ANOVA analysis results from SPSS

By the result analysis of variance (ANOVA), the null hypothesis could be accepted where there is no big difference between the two groups of OLS and GP because the significance is 0.974 less than $\alpha = 0.05$

8. Conclusion:

It was found that exploiting the goal programming method in the field of estimating the parametric coefficients is not less efficient than the well-known classical method (OLS). This conclusion is depending upon the obtained results from both methods. The table of variance analysis confirms this deduction. Moreover, the goal programming method has an advantage over the (OLS) begin softer and able to process more than one goal simultaneously. This advantage might make the statistician able to make the convenient decision to solve the problem of multiple linear regression in case of multiple goals.

References

[1] M. H. Ahmad and et al, "Comparing Least-Squares and Goal Programming Estimates of Linear Regression Parameters," MATEMATICA, vol. 21, no. 2, (2005) 101–112.

[2] A. Charne and W. Cooper, "Management Models and Industrial Application of Linear Programming," Manage. Sci., vol. 4, no. 1, (1957).

[3] A. Charne and W. Cooper, "Goal Programming and Multiple Objectives Optimizations.," Eur. J. Oper. Res., vol. 1, (1977) 39–54.

[4] L. AL.Degelae, "Building the workforce planning model in the General Establishment for Electricity Distribution of Baghdad," PhD thesis, Mustansiriya University, (1997).

[5] J. Ignazio, Goal Programming and Extensions, 2nd ed. Lexington Books, (1976).
[6] Y. Ijiri, “Management Goals and Accounting for Control,” Stud. Math. Manag. Econ., vol. 3, (1965).
[7] S. Kathem, “Comparisons between parameter estimations of linear regression model using OLS and programming goals,” J. of Administration and Econ. Al Muthana Univ., (2009).
[8] E. Tu, “A Comparison of Three Linear Programming Models for Computing Least Absolute Values Estimates,” J.OF Math .and Stat, vol. 34, (2005) 95–102.
[9] D. A. Wolfe, ”Nonparametrics: Statistical Methods Based on Ranks and Its Impact on the Field of Non-parametric Statistics,” Sel. Work. E. L. Lehmann, (2012) 1101–1110.