

Coefficient inequalities for a subclass of analytic functions associated with exponential function

G. SINGH AND G. SINGH

Abstract

This paper is concerned with the upper bound of various coefficient functionals for a certain subclass of analytic functions associated with exponential function in the open unit disc $E = \{z \in \mathbb{C} : |z| < 1\}$. This investigation will motivate other researchers to work in this direction.

Mathematics Subject Classification 2010: 30C45, 30C50

Keywords: Analytic functions, Subordination, Coefficient inequalities, Exponential function, Hankel determinant, Fekete-Szegő inequality.

1. INTRODUCTION

Geometric function theory is an active topic of current research. The most remarkable result in the theory of univalent functions was Bieberbach's conjecture, established by Bieberbach [4]. It states that, for $f \in \mathcal{S}$, $|a_n| \leq n$, $n = 2, 3, \dots$ and it remained as a challenge for the mathematicians for a long time. Finally, L. De-Branges [6], proved this conjecture in 1985. During the course of proving this conjecture, various results related to the coefficients came into existence which gave rise to some new subclasses of analytic functions.

Let us denote by \mathcal{A} , the class of analytic functions of the form $f(z) = z + \sum_{k=2}^{\infty} a_k z^k$, defined in the open unit disc $E = \{z \in \mathbb{C} : |z| < 1\}$ and normalized by the conditions $f(0) = f'(0) - 1 = 0$. \mathcal{S} denotes the subclass of \mathcal{A} , which consists of univalent functions in E .

Firstly, we discuss some fundamental classes of analytic functions, which play an important role in defining our main class:

$\mathcal{S}^* = \left\{ f : f \in \mathcal{A}, \operatorname{Re} \left(\frac{zf'(z)}{f(z)} \right) > 0, z \in E \right\}$, the class of starlike functions.

Reade [25] introduced the concept of close-to-star functions. The class of close-to-star functions generally denoted by $\mathcal{C}\mathcal{S}^*$, consists of functions $f \in \mathcal{A}$ such that $\operatorname{Re} \left(\frac{f(z)}{g(z)} \right) > 0$, $g \in \mathcal{S}^*$. For $g(z) = z$, MacGregor [18] studied the following subclass of

close-to-star functions:

$$\mathcal{R}' = \{f : f \in \mathcal{A}, \operatorname{Re} \left(\frac{f(z)}{z} \right) > 0, z \in E\}.$$

The class \mathcal{R} of bounded turning functions was introduced and studied by MacGregor [17] and is defined as

$$\mathcal{R} = \{f : f \in \mathcal{A}, \operatorname{Re}(f'(z)) > 0, z \in E\}.$$

Later on, Murugusundramurthi and Magesh [21] studied the following class:

$$\mathcal{R}(\alpha) = \left\{ f : f \in \mathcal{A}, \operatorname{Re} \left((1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) \right) > 0, z \in E \right\}.$$

Particularly, $\mathcal{R}(1) \equiv \mathcal{R}$ and $\mathcal{R}(0) \equiv \mathcal{R}'$.

An exponential function is a mathematical function, which is used in many real world situations. Mainly, it is used to find the exponential decay or exponential growth. The exponential function $\phi(z) = e^z$ has positive real part in E and $\phi(E) = \{z \in \mathbb{C} : |\log z| < 1\}$ is symmetric with respect to the real axis and starlike with respect to 1.

Let f and g be two analytic functions in E . We say that f is subordinate to g (denoted as $f \prec g$) if there exists a function w with $w(0) = 0$ and $|w(z)| < 1$ for $z \in E$ such that $f(z) = g(w(z))$. Further, if g is univalent in E , then this subordination leads to $f(0) = g(0)$ and $f(E) \subset g(E)$.

The concept of subordination played an important role in establishing many new classes of analytic functions. Mendiratta et al. [20] investigated the class $\mathcal{S}^*(e^z)$, the class of starlike functions associated with exponential function. Further Hai-Yan Zhang et al. [29] established the third Hankel determinant for the class $\mathcal{S}^*(e^z)$. Recently Ganesh et al. [9] studied the classes $\mathcal{S}_s^*(e^z)$ and $\mathcal{C}_s(e^z)$, the classes of starlike functions with respect to symmetric points and convex functions with respect to symmetric points associated with exponential function. Also Kumar and Sharma [12] studied the class $\mathcal{R}(e^z)$, the class of bounded turning functions associated with exponential function.

Getting motivated by these works, now we define the following class of analytic functions by subordinating to e^z .

DEFINITION 1. A function $f \in \mathcal{A}$ is said to be in the class $\mathcal{R}^\alpha(e^z)$ if it satisfies the condition

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) \prec e^z$$

or equivalently

$$\left| \log \left[(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) \right] \right| < 1.$$

We have the following observations:

- (i) $\mathcal{R}^0(e^z) \equiv \mathcal{R}'(e^z)$.
- (ii) $\mathcal{R}^1(e^z) \equiv \mathcal{R}(e^z)$.

For $q \geq 1$ and $n \geq 1$, Noonan and Thomas [22] introduced the q^{th} Hankel determinant as

$$H_q(n) = \begin{vmatrix} a_n & a_{n+1} & \dots & a_{n+q-1} \\ a_{n+1} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n+q-1} & \dots & \dots & a_{n+2q-2} \end{vmatrix}.$$

For $q = 2, n = 1$ and $a_1 = 1$, the Hankel determinant reduces to $H_2(1) = a_3 - a_2^2$, which is the well known Fekete-Szegő functional. Further, Fekete and Szegő [8] generalised the estimate of $|a_3 - \mu a_2^2|$ where μ is real and $f \in \mathcal{A}$. Also for $q = 2, n = 2$, the Hankel determinant takes the form of $H_2(2) = a_2 a_4 - a_3^2$, which is Hankel determinant of order 2.

The functional $J_{n,m}(f) = a_n a_m - a_{m+n-1}$, $n, m \in \mathbb{N} - \{1\}$, is known as generalized Zalcman functional. It was first investigated by Ma [16]. The functional $J_{2,3}(f) = a_2 a_3 - a_4$ is a specific case of the generalized Zalcman functional. The upper bound for the functional $J_{2,3}(f)$ over different subclasses of analytic functions was computed by various authors. It is very useful in establishing the bounds for the third Hankel determinant.

Furthermore, for $q = 3, n = 1$, the Hankel determinant $H_q(n)$ reduces to

$$H_3(1) = ||,$$

which is the third order Hankel determinant.

For $a_1 = 1$, $H_3(1)$ can be expanded as

$$H_3(1) = a_3(a_2 a_4 - a_3^2) - a_4(a_4 - a_2 a_3) + a_5(a_3 - a_2^2),$$

and after applying the triangle inequality, it yields

$$|H_3(1)| \leq |a_3||a_2a_4 - a_3^2| + |a_4||a_2a_3 - a_4| + |a_5||a_3 - a_2^2|. \quad (1)$$

Extensive work has been done on the estimation of second Hankel determinant by various authors including Noor [23], Ehrenborg [7], Layman [13], Singh [27], Mehrok and Singh [19] and Janteng et al. [10]. The estimation of third Hankel determinant is little bit complicated. Babalola [3] was the first researcher who successfully obtained the upper bound of third Hankel determinant for the classes of starlike functions, convex functions and the class of functions with bounded boundary rotation. Further a few researchers including Shanmugam et al. [26], Bucur et al. [5], Altinkaya and Yalcin [1], Singh and Singh [28] have been actively engaged in the study of third Hankel determinant for various subclasses of analytic functions.

In the present paper, we establish the upper bounds for the initial coefficients, Fekete-Szegö inequality, Zalcman functional, second Hankel determinant and third Hankel determinant, for the class $\mathcal{R}^\alpha(e^z)$. Also various known results follow as particular cases.

Let \mathcal{P} denote the class of analytic functions p of the form

$$p(z) = 1 + \sum_{k=1}^{\infty} p_k z^k,$$

whose real parts are positive in E .

In order to prove our main results, the following lemmas have been used:

LEMMA 1. [24; 11] If $p \in \mathcal{P}$, then

$$\begin{aligned} |p_k| &\leq 2, k \in \mathbb{N}, \\ \left| p_2 - \frac{p_1^2}{2} \right| &\leq 2 - \frac{|p_1|^2}{2}, \\ |p_{i+j} - \mu p_i p_j| &\leq 2, 0 \leq \mu \leq 1, \\ |p_{n+2k} - \lambda p_n p_k^2| &\leq 2(1 + 2\lambda), (\lambda \in \mathbb{R}), \\ |p_m p_n - p_k p_l| &\leq 4, (m+n = k+l; m, n \in \mathbb{N}), \end{aligned}$$

and for complex number ρ , we have

$$|p_2 - \rho p_1^2| \leq 2 \max\{1, |2\rho - 1|\}.$$

LEMMA 2. Let $p \in \mathcal{P}$, then

$$|Jp_1^3 - Kp_1p_2 + Lp_3| \leq 2|J| + 2|K - 2J| + 2|J - K + L|.$$

In particular, it is proved in [24] that

$$|p_1^3 - 2p_1p_2 + p_3| \leq 2.$$

LEMMA 3. [14; 15] If $p \in \mathcal{P}$, then

$$2p_2 = p_1^2 + (4 - p_1^2)x,$$

$$4p_3 = p_1^3 + 2p_1(4 - p_1^2)x - p_1(4 - p_1^2)x^2 + 2(4 - p_1^2)(1 - |x|^2)z,$$

for $|x| \leq 1$ and $|z| \leq 1$.

2. COEFFICIENT INEQUALITIES

THEOREM 1. If $f \in \mathcal{R}^\alpha(e^z)$, then

$$|a_2| \leq \frac{1}{1 + \alpha}, \tag{2}$$

$$|a_3| \leq \frac{1}{1 + 2\alpha}, \tag{3}$$

$$|a_4| \leq \frac{1}{1 + 3\alpha}, \tag{4}$$

and

$$|a_5| \leq \frac{37}{24(1 + 4\alpha)}. \tag{5}$$

The estimates are sharp.

PROOF. Since $f \in \mathcal{R}^\alpha(e^z)$, by the principle of subordination, we have

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) = e^{w(z)}. \tag{6}$$

Define $p(z) = \frac{1+w(z)}{1-w(z)} = 1 + p_1z + p_2z^2 + p_3z^3 + \dots$, which implies $w(z) = \frac{p(z)-1}{p(z)+1}$.

On expanding, we have

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) = 1 + (1 + \alpha)a_2z + (1 + 2\alpha)a_3z^2 + (1 + 3\alpha)a_4z^3 + (1 + 4\alpha)a_5z^4 + \dots \tag{7}$$

Also

$$e^{w(z)} = 1 + \frac{1}{2}p_1z + \left(\frac{p_2}{2} - \frac{p_1^2}{8}\right)z^2 + \left(\frac{p_1^3}{48} - \frac{p_1p_2}{4} + \frac{p_3}{2}\right)z^3 + \left(\frac{p_1^4}{384} + \frac{p_1^2p_2}{16} - \frac{p_3p_1}{4} - \frac{p_2^2}{8} + \frac{p_4}{2}\right)z^4 + \dots \quad (8)$$

Using (7) and (8), (6) yields

$$1 + (1 + \alpha)a_2z + (1 + 2\alpha)a_3z^2 + (1 + 3\alpha)a_4z^3 + (1 + 4\alpha)a_5z^4 + \dots = 1 + \frac{1}{2}p_1z + \left(\frac{p_2}{2} - \frac{p_1^2}{8}\right)z^2 + \left(\frac{p_1^3}{48} - \frac{p_1p_2}{4} + \frac{p_3}{2}\right)z^3 + \left(\frac{p_1^4}{384} + \frac{p_1^2p_2}{16} - \frac{p_3p_1}{4} - \frac{p_2^2}{8} + \frac{p_4}{2}\right)z^4 + \dots \quad (9)$$

Equating the coefficients of z , z^2 , z^3 and z^4 in (9) and on simplification, we obtain

$$a_2 = \frac{1}{2(1 + \alpha)}p_1, \quad (10)$$

$$a_3 = \frac{1}{1 + 2\alpha} \left[\frac{p_2}{2} - \frac{p_1^2}{8} \right], \quad (11)$$

$$a_4 = \frac{1}{48(1 + 3\alpha)} [p_1^3 - 12p_1p_2 + 24p_3], \quad (12)$$

and

$$a_5 = \frac{1}{(1 + 4\alpha)} \left[\frac{p_1^4}{384} - \frac{p_2^2}{8} - \frac{p_3p_1}{4} + \frac{p_1^2p_2}{16} + \frac{p_4}{2} \right]. \quad (13)$$

Using first inequality of Lemma 1 in (10), the result (2) is obvious.

From (11), we have

$$|a_3| = \frac{1}{2(1 + 2\alpha)} \left| p_2 - \frac{1}{4}p_1^2 \right|. \quad (14)$$

Using sixth inequality of Lemma 1 in (14), the result (3) can be easily obtained.

(12) can be expressed as

$$|a_4| = \frac{1}{48(1 + 3\alpha)} |p_1^3 - 12p_1p_2 + 24p_3|. \quad (15)$$

On applying Lemma 2 in (15), the result (4) is obvious.

Further, (13) can be re-written as

$$|a_5| = \frac{1}{2(1 + 4\alpha)} \left| \frac{1}{2} \left(p_4 - \frac{1}{2}p_2^2 \right) + \frac{1}{2}(p_4 - p_1p_3) + \frac{1}{8}p_1^2p_2 + \frac{1}{192}p_1^4 \right|. \quad (16)$$

On applying triangle inequality and using third inequality of Lemma 1, the result (5)

is obvious from (16).

The results (2), (3), (4) and (5) are sharp for the function f given by

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) = e^z.$$

On putting $\alpha = 0$, Theorem 1 yields the following result:

COROLLARY 1. If $f \in \mathcal{R}'(e^z)$, then

$$|a_2| \leq 1, |a_3| \leq 1, |a_4| \leq 1, |a_5| \leq \frac{37}{24}.$$

For $\alpha = 1$, Theorem 1 gives the following result due to Kumar and Sharma [12]:

COROLLARY 2. If $f \in \mathcal{R}(e^z)$, then

$$|a_2| \leq \frac{1}{2}, |a_3| \leq \frac{1}{3}, |a_4| \leq \frac{1}{4}, |a_5| \leq \frac{37}{120}.$$

THEOREM 2. If $f \in \mathcal{R}^\alpha(e^z)$, then

$$|a_3 - a_2^2| \leq \frac{1}{1 + 2\alpha}. \tag{17}$$

PROOF. From (10) and (11), we have

$$|a_3 - a_2^2| = \frac{1}{2(1 + 2\alpha)} \left| p_2 - \frac{\alpha^2 + 6\alpha + 3}{4(1 + \alpha)^2} p_1^2 \right|. \tag{18}$$

Using sixth inequality of Lemma 1, (18) takes the form

$$|a_3 - a_2^2| \leq \frac{1}{1 + 2\alpha} \max \left\{ 1, \frac{-1 + 2\alpha - \alpha^2}{2(1 + \alpha)^2} \right\}. \tag{19}$$

But $\frac{-1 + 2\alpha - \alpha^2}{2(1 + \alpha)^2} \leq 1$ for $0 \leq \alpha \leq 1$.

Hence, the result (17) is obvious from (19).

Substituting for $\alpha = 0$, Theorem 2 yields the following result:

COROLLARY 3. If $f \in \mathcal{R}'(e^z)$, then

$$|a_3 - a_2^2| \leq 1.$$

Putting $\alpha = 1$, Theorem 2 yields the following result due to Kumar and Sharma [12]:

COROLLARY 4. If $f \in \mathcal{R}(e^z)$, then

$$|a_3 - a_2^2| \leq \frac{1}{3}.$$

THEOREM 3. If $f \in \mathcal{R}^\alpha(e^z)$, then

$$|a_2a_3 - a_4| \leq \frac{1}{1+3\alpha}. \quad (20)$$

PROOF. Using (10), (11), (12) and after simplification, we have

$$|a_2a_3 - a_4| = \frac{1}{48(1+\alpha)(1+2\alpha)(1+3\alpha)} |(4+12\alpha+2\alpha^2)p_1^3 - (24+72\alpha+24\alpha^2)p_1p_2+24(1+\alpha)(1+2\alpha)p_3|. \quad (21)$$

On applying Lemma 2 in (21), it yields (20).

For $\alpha = 0$, the following result is a consequence of Theorem 3:

COROLLARY 5. If $f \in \mathcal{R}'(e^z)$, then

$$|a_2a_3 - a_4| \leq 1.$$

On putting $\alpha = 1$ in Theorem 3, we can obtain the following result due to Kumar and Sharma [12]:

COROLLARY 6. If $f \in \mathcal{R}(e^z)$, then

$$|a_2a_3 - a_4| \leq \frac{1}{4}.$$

THEOREM 4. If $f \in \mathcal{R}^\alpha(e^z)$, then

$$|a_2a_4 - a_3^2| \leq \frac{1}{(1+2\alpha)^2}. \quad (22)$$

The bound is sharp.

PROOF. Using (10), (11) and (12), we have

$$|a_2a_4 - a_3^2| = \frac{1}{192(1+\alpha)(1+2\alpha)^2(1+3\alpha)} |48(1+2\alpha)^2p_1p_3 - 24\alpha^2p_1^2p_2 + (-1-4\alpha-\alpha^2)p_1^4 - 48(1+\alpha)(1+3\alpha)p_2^2|.$$

Substituting for p_2 and p_3 from Lemma 3 and letting $p_1 = p$, we get

$$|a_2a_4 - a_3^2| = \frac{1}{192(1+\alpha)(1+2\alpha)^2(1+3\alpha)} \left| -(\alpha^2 + 4\alpha + 1)p^4 + 12\alpha^2 p^2(4 - p^2)x - 12(1 + 2\alpha)^2 p^2(4 - p^2)x^2 - 12(1 + \alpha)(1 + 3\alpha)(4 - p^2)^2 x^2 + 24(1 + 2\alpha)^2 p(4 - p^2)(1 - |x|^2)z \right|.$$

Since $|p| = |p_1| \leq 2$, we may assume that $p \in [0, 2]$. By using triangle inequality and $|z| \leq 1$ with $|x| = t \in [0, 1]$, we obtain

$$|a_2a_4 - a_3^2| \leq \frac{1}{192(1+\alpha)(1+2\alpha)^2(1+3\alpha)} \left[(\alpha^2 + 4\alpha + 1)p^4 + 12\alpha^2 p^2(4 - p^2)t + 12(1 + 2\alpha)^2 p^2(4 - p^2)t^2 + 12(1 + \alpha)(1 + 3\alpha)(4 - p^2)^2 t^2 + 24(1 + 2\alpha)^2 p(4 - p^2) - 24(1 + 2\alpha)^2 p(4 - p^2)t^2 \right] = F(p, t).$$

$$\frac{\partial F}{\partial t} = \frac{1}{192(1+\alpha)(1+2\alpha)^2(1+3\alpha)} \left[12\alpha^2 p^2(4 - p^2) + 24(4 - p^2)(2 - p)t[\alpha^2(6 - p) + 8\alpha + 2] \right] \geq 0,$$

and so $F(p, t)$ is an increasing function of t .

Therefore, $\max\{F(p, t)\} = F(p, 1) = \frac{1}{192(1+\alpha)(1+2\alpha)^2(1+3\alpha)} \left[(\alpha^2 + 4\alpha + 1)p^4 + 12\alpha^2 p^2(4 - p^2) + 12(1 + 2\alpha)^2 p^2(4 - p^2) + 12(1 + \alpha)(1 + 3\alpha)(4 - p^2)^2 \right] = H(p).$

$H'(p) = 0$ gives $p = 0$. Also $H''(p) < 0$ for $p = 0$.

This implies $\max\{H(p)\} = H(0) = \frac{1}{(1+2\alpha)^2}$, which proves (22).

The result is sharp for $p_1 = 0, p_2 = \pm 2$ and $p_3 = 0$.

Putting $\alpha = 0$, Theorem 4 gives the following result:

COROLLARY 7. If $f \in \mathcal{H}(e^z)$, then

$$|a_2a_4 - a_3^2| \leq 1.$$

Substituting for $\alpha = 1$ in Theorem 4, the following result due to Kumar and Sharma [12], is obvious:

COROLLARY 8. If $f \in \mathcal{H}(e^z)$, then

$$|a_2a_4 - a_3^2| \leq \frac{1}{9}.$$

THEOREM 5. If $f \in \mathcal{H}^\alpha(e^z)$, then

$$|H_3(1)| \leq \frac{85 + 850\alpha + 3025\alpha^2 + 4428\alpha^3 + 2100\alpha^4}{24(1 + 2\alpha)^3(1 + 3\alpha)^2(1 + 4\alpha)}. \tag{23}$$

PROOF. By using (3), (4), (5), (17), (20) and (22) in (1), the result (23) can be easily obtained.

For $\alpha = 0$, Theorem 5 yields the following result:

COROLLARY 9. If $f \in \mathcal{R}'(e^z)$, then

$$|H_3(1)| \leq \frac{85}{24}.$$

For $\alpha = 1$, Theorem 5 yields the following result:

COROLLARY 10. If $f \in \mathcal{R}(e^z)$, then

$$|H_3(1)| \leq \frac{437}{2160}.$$

3. BOUNDS OF $|H_3(1)|$ FOR TWO-FOLD AND THREE-FOLD SYMMETRIC FUNCTIONS

A function f is said to be n -fold symmetric if it satisfies the following condition:

$$f(\xi z) = \xi f(z)$$

where $\xi = e^{\frac{2\pi i}{n}}$ and $z \in E$.

By $S^{(n)}$, we denote the set of all n -fold symmetric functions which belong to the class S .

The n -fold univalent function has the following Taylor-Maclaurin series:

$$f(z) = z + \sum_{k=1}^{\infty} a_{nk+1} z^{nk+1}. \quad (24)$$

An analytic function f of the form (24) belongs to the family $\mathcal{R}_{car}^{\alpha(n)}$ if and only if

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) = e^{\left(\frac{p(z)-1}{p(z)+1}\right)}, p \in \mathcal{P}^{(n)},$$

where

$$\mathcal{P}^{(n)} = \left\{ p \in \mathcal{P} : p(z) = 1 + \sum_{k=1}^{\infty} p_{nk} z^{nk}, z \in E \right\}. \quad (25)$$

THEOREM 6. If $f \in \mathcal{R}^{\alpha(2)}(e^z)$, then

$$|H_3(1)| \leq \frac{1}{(1+2\alpha)(1+4\alpha)}. \quad (26)$$

PROOF. If $f \in \mathcal{R}^{\alpha(2)}(e^z)$, so there exists a function $p \in \mathcal{P}^{(2)}$ such that

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) = e^{\left(\frac{p(z)-1}{p(z)+1}\right)}. \tag{27}$$

Using (24) and (25) for $n = 2$, (27) yields

$$a_3 = \frac{1}{2(1+2\alpha)} p_2, \tag{28}$$

$$a_5 = \frac{1}{2(1+4\alpha)} \left(p_4 - \frac{1}{4} p_2^2 \right). \tag{29}$$

Also

$$H_3(1) = a_3 a_5 - a_3^3. \tag{30}$$

Using (28) and (29) in (30), it yields

$$H_3(1) = \frac{1}{8(1+2\alpha)(1+4\alpha)} p_2 \left[p_4 - \frac{3(1+2\alpha)^2 + 8(1+4\alpha)}{12(1+2\alpha)^2} p_2^2 \right]. \tag{31}$$

On applying triangle inequality and using fourth inequality of Lemma 1, we can easily get the result (26).

Putting $\alpha = 0$, the following result can be easily obtained from Theorem 6:

COROLLARY 11. If $f \in \mathcal{R}^{(2)}(e^z)$, then

$$|H_3(1)| \leq 1.$$

For $\alpha = 1$, Theorem 6 agrees with the following result:

COROLLARY 12. If $f \in \mathcal{R}^{(2)}(e^z)$, then

$$|H_3(1)| \leq \frac{1}{15}.$$

THEOREM 7. If $f \in \mathcal{R}^{\alpha(3)}(e^z)$, then

$$|H_3(1)| \leq \frac{1}{(1+3\alpha)^2}. \tag{32}$$

PROOF. If $f \in \mathcal{R}^{\alpha(3)}(e^z)$, so there exists a function $p \in \mathcal{P}^{(3)}$ such that

$$(1 - \alpha) \frac{f(z)}{z} + \alpha f'(z) = e^{\left(\frac{p(z)-1}{p(z)+1}\right)}. \tag{33}$$

Using (24) and (25) for $n = 3$, (33) gives

$$a_4 = \frac{1}{2(1+3\alpha)} p_3. \quad (34)$$

Also

$$H_3(1) = -a_4^2. \quad (35)$$

Using (34) in (35), it yields

$$H_3(1) = -\frac{1}{4(1+3\alpha)^2} p_3^2. \quad (36)$$

On applying triangle inequality and using first inequality of Lemma 1, (32) can be easily obtained.

For $\alpha = 0$, Theorem 7 yields the following result:

COROLLARY 13. If $f \in \mathcal{R}^{(3)}(e^z)$, then

$$|H_3(1)| \leq 1.$$

For $\alpha = 1$, Theorem 7 yields the following result:

COROLLARY 14. If $f \in \mathcal{R}^{(3)}(e^z)$, then

$$|H_3(1)| \leq \frac{1}{16}.$$

CONCLUSION

This paper is concerned with the upper bound of third Hankel determinant for a subclass of analytic functions, associated with exponential function. The results established here are very interesting and will motivate other researchers in this field to work on the similar classes by associating to other standard functions.

REFERENCES

- S. Altinkaya and S. Yalcin, Third Hankel determinant for Bazilevic functions, *Adv. Math., Scientific Journal*, 5(2)(2016), 91-96.
- M. Arif, M. Raza, H. Tang, S. Hussain and H. Khan, Hankel determinant of order three for familiar subsets of analytic functions related with sine function, *Open Math.*, 17(2019), 1615-1630.
- K. O. Babalola, On $H_3(1)$ Hankel determinant for some classes of univalent functions, *Ineq. Th. Appl.*, 6(2010), 1-7.
- L. Bieberbach, Über die koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln, *Sitzungsberichte Preussische Akademie der Wissenschaften*, 138(1916), 940-955.
- R. Bucur, D. Breaz and L. Georgescu, Third Hankel determinant for a class of analytic functions with respect to symmetric points, *Acta Univ. Apulensis*, 42(2015), 79-86.
- L. De-Branges, A proof of the Bieberbach conjecture, *Acta Math.*, 154(1985), 137-152.
- R. Ehrenborg, The Hankel determinant of exponential polynomials, *Amer. Math. Monthly*, 107(2000), 557-560.
- M. Fekete and G. Szegő, Eine Bemerkung über ungerade schlichte Funktionen, *J. Lond. Math. Soc.*, 8(1933), 85-89.
- K. Ganesh, R. B. Sharma and K. R. Laxmi, Third Hankel determinant for a class of functions with respect to symmetric points associated with exponential function, *WSEAS Trans. Math.*, 19(2020), 133-138.
- A. Janteng, S. A. Halim and M. Darus, Hankel determinant for starlike and convex functions, *Int. J. Math. Anal.*, 1(13)(2007), 619-625.
- S. R. Keogh and E. P. Merkes, A coefficient inequality for certain subclasses of analytic functions, *Proc. Amer. Math. Soc.*, 20(1969), 8-12.
- V. S. Kumar and R. B. Sharma, A study on Zalcman conjecture and third Hankel determinant, *J. Phy.: Conference Series*, 1597(2020).
- J. W. Layman, The Hankel transform and some of its properties, *J. Int. Seq.*, 4(2001), 1-11.
- R. J. Libera and E. J. Zlotkiewicz, Early coefficients of the inverse of a regular convex function, *Proc. Amer. Math. Soc.*, 85(1982), 225-230.
- R. J. Libera and E. J. Zlotkiewicz, Coefficient bounds for the inverse of a function with derivative in \mathcal{P} , *Proc. Amer. Math. Soc.*, 87(1983), 251-257.
- W. Ma, Generalized Zalcman conjecture for starlike and typically real functions, *J. Math. Anal. Appl.*, 234(1999), 328-329.
- T. H. MacGregor, Functions whose derivative has a positive real part, *Trans. Amer. Math. Soc.*, 104(1962), 532-537.
- T. H. MacGregor, The radius of univalence of certain analytic functions, *Proc. Amer. Math. Soc.*, 14(1963), 514-520.
- B. S. Mehrook and G. Singh, Estimate of second Hankel determinant for certain classes of analytic functions, *Scientia Magna*, 8(3)(2012), 85-94.
- R. Mendiratta, S. Nagpal and V. Ravichandran, On a subclass of strongly starlike functions associated with exponential function, *Bull. Malays. Math. Sci. Soc.*, 38(1)(2015), 365-386.

- G. Murugusundramurthi and N. Magesh, Coefficient inequalities for certain classes of analytic functions associated with Hankel determinant, *Bull. Math. Anal. Appl.*, 1(3)(2009), 85-89.
- J. W. Noonan and D. K. Thomas, On the second Hankel determinant of a really mean p -valent functions, *Trans. Amer. Math. Soc.*, 223(2)(1976), 337-346.
- K. I. Noor, Hankel determinant problem for the class of functions with bounded boundary rotation, *Rev. Roum. Math. Pures Et Appl.*, 28(8)(1983), 731-739.
- Ch. Pommerenke, Univalent functions, *Math. Lehrbuecher, Vandenhoeck and Ruprecht, Göttingen*, 1975.
- M. O. Reade, On close-to-convex univalent functions, *Michigan Math. J.*, 3(1955-56), 59-62.
- G. Shanmugam, B. A. Stephen and K. O. Babalola, Third Hankel determinant for α starlike functions, *Gulf J. Math.*, 2(2)(2014), 107-113.
- G. Singh, Hankel determinant for a new subclass of analytic functions, *Scientia Magna*, 8(4)(2012), 61-65.
- G. Singh and G. Singh, On third Hankel determinant for a subclass of analytic functions, *Open Sci. J. Math. Appl.*, 3(6)(2015), 172-175.
- H. Y. Zhang, H. Tang and X. M. Niu, Third order Hankel determinant for certain class of analytic functions related with exponential function, *Symmetry*, 10(501)(2018), doi:10.3390/sym.10100501.

Gagandeep Singh
Department of Mathematics
Khalsa College, Amritsar-143002, Punjab, India
Email: kamboj.gagandeep@yahoo.in

Gurcharanjit Singh
Department of Mathematics
G.N.D.U. College, Chungh (Punjab), India.

Trichotomization with two cutoff values using Kruskal-Wallis test by minimum P -value approach

T. OGURA AND C. SHIRAISHI

Abstract

In clinical trials, age is often converted to binary data by the cutoff value. However, when looking at a scatter plot for a group of patients whose age is larger than or equal to the cutoff value, age and outcome may not be related. If the group whose age is greater than or equal to the cutoff value is further divided into two groups, the older of the two groups may appear to be at lower risk. In this case, it may be necessary to further divide the group of patients whose age is greater than or equal to the cutoff value into two groups. This study provides a method for determining which of the two or three groups is the optimal division. The following two methods are used to divide the data. The existing method, the Wilcoxon-Mann-Whitney test by minimum P -value approach, divides data into two groups by one cutoff value. A new method, the Kruskal-Wallis test by minimum P -value approach, divides data into three groups by two cutoff values. Of the two tests, the one with the smaller P -value is used. Because this was a new decision procedure, it was tested using Monte Carlo simulations (MCSs) before application to the available COVID-19 data. The MCS results showed that this method performs well. In the COVID-19 data, it was optimal to divide into three groups by two cutoff values of 60 and 70 years old. By looking at COVID-19 data divided into three groups according to the two cutoff values, it was confirmed that each group had different features. We provided the R code that can be used to replicate the results of this manuscript. Another practical example can be performed by replacing x and y with appropriate ones.

Mathematics Subject Classification 2010: 62P10

General Terms: Algorithms

Keywords: COVID-19 data, cutoff value, Kruskal-Wallis test, minimum P -value approach, Wilcoxon-Mann-Whitney test.

1. INTRODUCTION

In the medical field, age is frequently converted to binary data for analysis, depending on whether it is less than the cutoff value or greater than or equal to the cutoff value [Mirkes et al. 2016]. The risk generally rises with age. As a result, the group of patients whose age is less than the cutoff value being interpreted as low risk and the group of patients whose age is greater than or equal to the cutoff value being interpreted as high risk. However, in clinical trials, if the group whose age is greater than or equal to the cutoff value is further divided into two groups, the older of the two groups may appear to be at lower risk. This could be due to several factors. One probable cause is that due to inclusion and exclusion criteria, patients at increased

risk as they grow older may be excluded from clinical trials, and participants may be limited to low risk patients. Age and outcome may not be correlated in clinical trials, even though age and outcome are correlated in the clinical setting. As a result, in clinical trials, it may be better to analyze three groups rather than two groups.

When dependent and independent variables are binary data and quantitative data, respectively, the receiver operating characteristic curve is used to find the optimal cutoff value for predicting outcome [Greiner et al. 2000; Zou et al. 2007]. When both dependent and independent variables are quantitative data, the piecewise linear regression analysis improves the accuracy of outcome prediction by changing the linear regression line at the cutoff value [Nakamura 1986; Vieth 1989]. However, the piecewise linear regression analysis may not yield the optimal cutoff value when the data follow a non-normal distribution. As the method for determining the optimal cutoff value for such data, Ogura and Shiraishi [2022] showed the method of performing the Wilcoxon-Mann-Whitney (WMW) test on all potential cutoff values and adopting the cutoff value that minimizes the P -value (called minimum P -value approach). The minimum P -value approach strategy performed well in log-rank and chi-square tests [Altman et al. 1994; Mazumdar and Glassman 2000; Liu et al. 2020], and it was thought to apply to other tests. Another known method is to test all potential cutoff values and adopt the cutoff value that maximizes the test statistic [?; ?; ?]. This method is called the maximum test statistic approach in this manuscript. The minimum P -value approach and maximum test statistic approach are based on similar ideas and therefore often give the same cutoff values. However, in tests where the P -value is calculated from the test statistic and degrees of freedom, the different degrees of freedom give different P -value for the same value of the test statistic. In this case, the minimum P -value approach and maximum test statistic approach may give different cutoff values. Because the minimum P -value approach is tested at all potential cutoff values, it raises the issue of multiple tests. To address this issue, various methods of adjusted P -value for multiple tests were proposed [?; ?; ?; ?]. In maximum test statistic approach, the adjusted P -value is considered appropriate instead of the standard P -value [?; ?].

In this manuscript, the trichotomization is illustrated on an analysis of data related to COVID-19. The first COVID-19 patients were reported in Wuhan, China, in 2019, and the virus has since spread worldwide [World Health Organization 2020; 2022]. The data obtained under the circumstances of COVID-19 are often different from the conventional ones. Researchers have been researching suitable analysis methods for

the data [Ünözkan et al. 2020]. Reznikov et al. [2022] divided the COVID-19 data into categories based on patient age (30 years or younger, 31–60 years, 61 years or older). They obtained useful results by dividing the COVID-19 data into three groups. Ogura and Shiraishi [2022] searched for the one optimal cutoff value using the supplementary data of Hogan II et al. [2020] when dependent and independent variables were age and days to discharge in COVID-19 data. At the one cutoff value of 59.5 years old, they were able to divide the data into two groups with differing characteristics. However, the older group tends to have shorter days to discharge as they get older. Therefore, we investigate whether the COVID-19 data should be divided into two groups or three groups. We use the Kruskal-Wallis (KW) test by the minimum P -value approach to divide into three groups. The KW test is a nonparametric alternative of one-way analysis of variance and its properties were investigated by [?]. As the method to divide into two groups, we adopt the one cutoff value for the WMW test by the minimum P -value approach. As the method to divide into three groups, we adopt the two cutoff values for the KW test by the minimum P -value approach. We adopt the cutoff value(s) corresponding to the smaller P -value of these two tests. Ogura and Shiraishi [2022] validated the WMW test by the minimum P -value approach. Because the KW test by the minimum P -value approach is a new method, using Monte Carlo simulations (MCSs) at diverse situations, we tested the performance of the two cutoff values for the KW test by the minimum P -value approach. Using the COVID-19 data, we then demonstrated that it is preferable to divide data into three groups rather than two, with the two cutoff values being the ages of 60 and 70 years old.

Although it is also considered to use the maximum test statistic approach, the test statistics for the WMW and KW tests cannot be directly compared. In contrast, the P -values for the WMW and KW tests can be directly compared. Because the WMW and KW tests are performed on one dataset, the issue of multiple tests is usually larger in this study. The methods of the adjusted P -value in the minimum P -value approach were often validated in one type of test but not in two. Therefore, this study uses the widely available Bonferroni method [?]. The Bonferroni method is defined as the significance level divided by the number of tests performed. The adjusted P -value, obtained by multiplying the P -value by the number of tests performed, is sometimes used, which is essentially the same. We use the Bonferroni method as defined. Because this study uses two types of tests, the tests are performed many times. In addition, the potential combinations of two cutoff values are much more than one cutoff value. Therefore, it may be very difficult for the P -value to be below

the significance level adjusted by the Bonferroni method.

In Section 2, we describe the existing method of the one cutoff value for the WMW test and a new method of the two cutoff values for the KW test by the minimum P -value approach. In Section 3, we test the performance of the cutoff value(s) using MCSs. In Section 4, using COVID-19 data, we present a judgment of the number of cutoff values and an attempt to determine to the cutoff value(s). Finally, we conclude the research in Section 5.

2. OPTIMAL CUTOFF VALUE(S)

Let $(x, y) = \{(x_1, y_1), \dots, (x_n, y_n)\}$ be two dimensional random vectors of sample size $n \geq 3$, where x and y are independent and dependent variables, respectively. Furthermore, it is assumed that x is age and the value of each x_i ($i = 1, \dots, n$) is an integer in this manuscript. Let $x_{(i)}$ denote the i -th order statistics, $x_{(1)} \leq \dots \leq x_{(n)}$. Section 2.1 presents the existing method to divide into two groups by the value of $x_{(i)}$. Section 2.2 presents a new method to divide into three groups by the value of $x_{(i)}$. Section 2.3 explains whether dividing the dataset into two or three groups is optimal.

2.1. One cutoff value for the WMW test

Ogura and Shiraishi [2022] demonstrated the methodology of one cutoff value using the WMW test by the minimum P -value approach. Because x is age and the value of each $x_{(i)}$ is an integer, it is illustrated below in a slightly different way from their paper. Let the cutoff value be c^W . The potential cutoff value is written as $c_{(j)}^W = x_{(j)}$, $j = 2, \dots, n$. The data are divided into two groups, $\{(x_{(1)}, y_{(1)}), \dots, (x_{(j-1)}, y_{(j-1)})\}$ and $\{(x_{(j)}, y_{(j)}), \dots, (x_{(n)}, y_{(n)})\}$, depending on whether $x_{(i)} < c_{(j)}^W$ or $c_{(j)}^W \leq x_{(i)}$, where $y_{(i)}$ is the data paired with $x_{(i)}$ ($y_{(i)}$ is not the order statistic of y_i). We perform the WMW test between $\{y_{(1)}, \dots, y_{(j-1)}\}$ and $\{y_{(j)}, \dots, y_{(n)}\}$ in sequence from $j = 2$ to n , and the P -value is written as $P_{(j)}^W$. The optimal cutoff value is $c^W = c_{(j)}^{W, \min}$ corresponding to $P_{(j)}^{W, \min} = \min(P_{(2)}^W, \dots, P_{(n)}^W)$. We utilize the one cutoff value where each group has ten or more patients in this manuscript since dividing by the one cutoff value has no benefit when the sample size of one group is small.

2.2. Two cutoff values for the KW test

Let the two cutoff values be c^{K_1} and c^{K_2} ($c^{K_1} < c^{K_2}$). The two potential cutoff values are written as $c_{(k)}^{K_1} = x_{(k)}$ and $c_{(l)}^{K_2} = x_{(l)}$, $k = 2, \dots, n-1$, $l = k+1, \dots, n$. The data are divided into three groups, $\{(x_{(1)}, y_{(1)}), \dots, (x_{(k-1)}, y_{(k-1)})\}$, $\{(x_{(k)}, y_{(k)}), \dots, (x_{(l-1)}, y_{(l-1)})\}$, and $\{(x_{(l)}, y_{(l)}), \dots, (x_{(n)}, y_{(n)})\}$, depending on whether $x_{(i)} < c_{(k)}^{K_1}$, $c_{(k)}^{K_1} \leq x_{(i)} < c_{(l)}^{K_2}$, or $c_{(l)}^{K_2} \leq$

$x_{(i)}$, where $y_{(i)}$ is the data paired with $x_{(i)}$ ($y_{(i)}$ is not the order statistic of y_i). We perform the KW test between $\{y_{(1)}, \dots, y_{(k-1)}\}$, $\{y_{(k)}, \dots, y_{(l-1)}\}$ and $\{y_{(l)}, \dots, y_{(n)}\}$ in sequence from $k = 2$ to $n - 1$ and $l = k + 1$ to n , and the P -value is written as $P_{(k,l)}^K$. The two optimal cutoff values are $c^{K_1} = c_{(k)}^{K_1 \cdot \min}$ and $c^{K_2} = c_{(l)}^{K_2 \cdot \min}$ corresponding to $P_{(k,l)}^{K \cdot \min} = \min(P_{(2,3)}^K, \dots, P_{(n-1,n)}^K)$. We utilize the two cutoff values where each group has ten or more patients in this manuscript since dividing by the two cutoff values has no benefit when the sample size of one group is small.

2.3. Choice of one cutoff value or two cutoff values

We compare $P_{(j)}^{W \cdot \min}$ of the minimum P -value using the WMW test with the $P_{(k,l)}^{K \cdot \min}$ of the minimum P -value using the KW test. When $P_{(j)}^{W \cdot \min} \leq P_{(k,l)}^{K \cdot \min}$, the data are divided into two groups by the one cutoff value c^W of the WMW test. When $P_{(j)}^{W \cdot \min} > P_{(k,l)}^{K \cdot \min}$, the data is divided into three groups by the two cutoff values c^{K_1} and c^{K_2} of the KW test.

3. MONTE CARLO SIMULATIONS

We test the effectiveness of the cutoff value(s) using MCSs. We set the number of population cutoff values as one and two in Sections 3.1 and 3.2, respectively. In Section 3.1, when the one population cutoff value is set to 50, we verify that the proportion of $P_{(j)}^{W \cdot \min} \leq P_{(k,l)}^{K \cdot \min}$ and the performance of the one cutoff value calculated by the WMW test are high. In Section 3.2, when the two population cutoff values are set to 50 and 70, we verify that the proportion of $P_{(j)}^{W \cdot \min} > P_{(k,l)}^{K \cdot \min}$ and the performance of the two cutoff values calculated by the KW test are high. When $x_{(i)} < 70$, the risks in Sections 3.1 and 3.2 have the same settings. In detail, when $x_i < 50$, the risk is set low, and when $50 \leq x_i < 70$, the risk is set high. In Section 3.1, the risks are the same for $50 \leq x_i < 70$ and $70 \leq x_i$. In Section 3.2, the risk of $70 \leq x$ is set to be less than the risk of $50 \leq x_i < 70$. These settings confirm that after dividing the age into two groups by one cutoff value, it is possible to correctly identify whether the elderly group should be further divided into two groups or not. Because age and days to discharge in the COVID-19 data are both integers, x_i and y_i are rounded to the nearest whole number in MCSs. The sample size is set to $n = 80, 100, 120$. The replication size used in this study is 100,000. We use the software R version 4.1.1 [R Core Team 2021] for the MCSs. The MCSs are carried out according to the steps below.

- (1) Generate random samples $\{x_1, \dots, x_n\}$ from distribution in Table I or Table IV, and round them to integers.
- (2) In Section 3.1, generate random samples $\{y_1, \dots, y_n\}$ from distribution

corresponding to $x_i < 50$ or $50 \leq x_i$ in Table I, and round them to integers.

- (3) In Section 3.2, generate random samples $\{y_1, \dots, y_n\}$ from distribution corresponding to $x_i < 50$, $50 \leq x_i < 70$, or $70 \leq x_i$ in Table IV, and round them to integers.
- (4) Combine $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\}$ into two-dimensional random vectors $\{(x_1, y_1), \dots, (x_n, y_n)\}$.
- (5) Sort $\{x_1, \dots, x_n\}$ in ascending order, $x_{(1)} \leq \dots \leq x_{(n)}$.
- (6) Set one potential cutoff value $c_{(j)}^W = x_{(j)}$, $j = 11, \dots, n-9$.
- (7) Divide into two groups, $\{(x_{(1)}, y_{(1)}), \dots, (x_{(j-1)}, y_{(j-1)})\}$ and $\{(x_{(j)}, y_{(j)}), \dots, (x_{(n)}, y_{(n)})\}$, depending on whether $x_{(i)} < c_{(j)}^W$ or $c_{(j)}^W \leq x_{(i)}$.
- (8) Perform the WMW test between two groups for each $c_{(j)}^W$ and express the P -value as $P_{(j)}^W$.
- (9) Repeat steps 6–8 from $j = 11, \dots, n-9$.
- (10) Decide optimal cutoff value $c^W = c_{(j)}^{W-\min}$ that satisfies $P_{(j)}^{W-\min} = \min(P_{(11)}^W, \dots, P_{(n-9)}^W)$.
- (11) Set two potential cutoff values $c_{(k)}^{K_1} = x_{(k)}$ and $c_{(l)}^{K_2} = x_{(l)}$, $k = 11, \dots, n-19$, $l = k+10, \dots, n-9$.
- (12) Divide into three groups, $\{(x_{(1)}, y_{(1)}), \dots, (x_{(k-1)}, y_{(k-1)})\}$, $\{(x_{(k)}, y_{(k)}), \dots, (x_{(l-1)}, y_{(l-1)})\}$ and $\{(x_{(l)}, y_{(l)}), \dots, (x_{(n)}, y_{(n)})\}$, depending on whether $x_{(i)} < c_{(k)}^{K_1}$, $c_{(k)}^{K_1} \leq x_{(i)} < c_{(l)}^{K_2}$ or $c_{(l)}^{K_2} \leq x_{(i)}$.
- (13) Perform the KW test between three groups for each $c_{(k)}^{K_1}$ and $c_{(l)}^{K_2}$ and express the P -value as $P_{(k,l)}^K$.
- (14) Repeat steps 11–13 from $k = 11, \dots, n-19$ and $l = k+10, \dots, n-9$.
- (15) Decide optimal two cutoff values $c^{K_1} = c_{(k)}^{K_1-\min}$ and $c^{K_2} = c_{(l)}^{K_2-\min}$ that satisfies $P_{(k,l)}^{K-\min} = \min(P_{(11,21)}^K, \dots, P_{(n-19,n-9)}^K)$.
- (16) Compare $P_{(j)}^{W-\min}$ and $P_{(k,l)}^{K-\min}$ and use the smaller one.
- (17) Independently, repeat steps 1–16 100,000 times.
- (18) Calculate the proportion of the number of the chosen cutoff values and the proportion of the cutoff value(s) in range.

3.1. One cutoff value

Table I presents the simulation settings for Patterns 1–9. Random samples of x_i and y_i are both generated from the three-parameter gamma distribution $\text{Ga}(\alpha, \beta, \gamma)$, where

Table I. Parameters of gamma distribution in Patterns 1–9.

Pattern	γ_x	α_y	β_y	γ_y
1	25	3.5	10	10
2	30	3.5	10	10
3	35	3.5	10	10
4	25	1.5	20	20
5	30	1.5	20	20
6	35	1.5	20	20
7	25	1.5	10	30
8	30	1.5	10	30
9	35	1.5	10	30

Table II. Proportion of the chosen one cutoff value in Section 3.1

Pattern	$n = 80$	$n = 100$	$n = 120$
1	88.1%	82.8%	83.0%
2	80.0%	78.1%	77.8%
3	69.3%	69.3%	69.5%
4	91.2%	86.8%	86.9%
5	83.1%	82.0%	82.0%
6	72.5%	72.7%	73.1%
7	94.7%	91.8%	91.8%
8	88.6%	87.2%	87.1%
9	77.9%	78.2%	78.8%

α , β , and γ are the shape, scale, and location parameters, respectively. First, x_i is generated from $\text{Ga}(2, 20, \gamma_x)$. Then, y_i is generated from $\text{Ga}(1.5, 10, 10)$ (if $x_i < 50$) or $\text{Ga}(\alpha_y, \beta_y, \gamma_y)$ (if $50 \leq x_i$).

Table II indicates the proportion of the chosen one cutoff value. When the number of population cutoff values is set to one, there is a high proportion that the number of cutoff values is appropriately determined to be one. Table III presents the proportion of the cutoff value that falls into five ranges ($49 \leq c^W \leq 51$, $48 \leq c^W \leq 52$, $47 \leq c^W \leq 53$, $46 \leq c^W \leq 54$, $45 \leq c^W \leq 55$). When $n = 80$, the proportion of the cutoff value was greater than 90% in the range of $45 \leq c^W \leq 55$. When $n = 100$, the proportion of the cutoff value was greater than 90% in the range of $47 \leq c^W \leq 53$. As a result, it was determined that the number of cutoff values was correctly judged and that the cutoff value was calculated close to the population cutoff value.

3.2. Two cutoff values

Table IV presents the simulation settings for Patterns 10–18. First, x_i is generated from the three-parameter gamma distribution $\text{Ga}(2, 20, \gamma_x)$. Then, y_i is generated from $\text{Ga}(1.5, 10, 10)$ (if $x_i < 50$), $\text{Ga}(\alpha_{y_1}, \beta_{y_1}, \gamma_{y_1})$ (if $50 \leq x_i < 70$), or $\text{Ga}(\alpha_{y_2}, \beta_{y_2}, \gamma_{y_2})$ (if $70 \leq x_i$).

Table III. Proportion of one cutoff value in range in Section 3.1

<i>n</i>	Pattern	49-51	48-52	47-53	46-54	45-55
80	1	65.9%	78.9%	86.4%	91.0%	93.9%
	2	67.8%	80.2%	86.9%	90.9%	93.5%
	3	64.7%	77.6%	84.4%	88.1%	90.4%
	4	68.9%	82.1%	89.1%	93.2%	95.6%
	5	71.7%	83.7%	89.8%	93.2%	95.3%
	6	67.9%	80.5%	86.9%	90.3%	92.3%
	7	71.7%	84.5%	91.3%	95.0%	97.1%
	8	74.9%	86.4%	92.0%	95.0%	96.8%
	9	71.5%	83.9%	89.9%	93.0%	94.7%
100	1	76.5%	87.6%	92.9%	95.7%	97.3%
	2	76.3%	87.2%	92.7%	95.5%	97.1%
	3	73.8%	85.2%	90.8%	93.7%	95.4%
	4	80.9%	90.9%	95.3%	97.3%	98.4%
	5	80.4%	90.4%	94.9%	97.1%	98.2%
	6	77.9%	88.5%	93.3%	95.6%	96.8%
	7	84.7%	93.5%	97.1%	98.6%	99.2%
	8	83.6%	92.6%	96.3%	98.2%	99.0%
	9	80.6%	90.9%	95.1%	97.1%	98.1%
120	1	81.6%	91.1%	95.3%	97.3%	98.4%
	2	81.4%	90.8%	94.9%	97.0%	98.2%
	3	78.9%	89.0%	93.6%	95.9%	97.2%
	4	85.5%	93.8%	97.1%	98.5%	99.2%
	5	85.0%	93.3%	96.8%	98.3%	99.0%
	6	82.6%	91.7%	95.5%	97.4%	98.3%
	7	88.9%	95.8%	98.2%	99.2%	99.6%
	8	88.0%	95.2%	97.9%	99.0%	99.5%
	9	85.2%	93.6%	97.0%	98.4%	99.1%

Table IV. Parameters of gamma distribution in Patterns 10–18.

Pattern	γ_k	α_{y_1}	β_{y_1}	γ_{y_1}	α_{y_2}	β_{y_2}	γ_{y_2}
10	25	3.5	10	10	2.5	10	10
11	30	3.5	10	10	2.5	10	10
12	35	3.5	10	10	2.5	10	10
13	25	1.5	20	20	1.5	15	15
14	30	1.5	20	20	1.5	15	15
15	35	1.5	20	20	1.5	15	15
16	25	1.5	10	30	1.5	10	20
17	30	1.5	10	30	1.5	10	20
18	35	1.5	10	30	1.5	10	20

Table V indicates the proportion of the chosen two cutoff values. When the number of population cutoff values is set to two, there is a high proportion that the number of cutoff values is appropriately determined to be two. As *n* increases, the proportion of correct judgments increases. Table VI presents the proportion of two cutoff values that falls into each of five ranges ($49 \leq c^{K_1} \leq 51$, $48 \leq c^{K_1} \leq 52$, $47 \leq c^{K_1} \leq 53$, $46 \leq c^{K_1} \leq 54$, $45 \leq c^{K_1} \leq 55$, $69 \leq c^{K_2} \leq 71$, $68 \leq c^{K_2} \leq 72$, $67 \leq c^{K_2} \leq 73$, $66 \leq c^{K_2} \leq 74$, $65 \leq c^{K_2} \leq 75$).

Table V. Proportion of the chosen two cutoff values in Section 3.2

Pattern	$n = 80$	$n = 100$	$n = 120$
10	70.6%	81.2%	85.7%
11	82.9%	87.7%	91.1%
12	89.3%	93.1%	95.3%
13	66.8%	80.1%	85.0%
14	81.9%	87.4%	91.0%
15	89.5%	93.2%	95.3%
16	82.4%	95.6%	97.9%
17	96.1%	98.5%	99.4%
18	98.2%	99.5%	99.9%

When $n = 80$, the cutoff value of c^{K_1} was greater than 90% in the range of $|c^{K_1} - 50| \leq 5$, except for Pattern 10. The cutoff value of c^{K_1} was greater than 90% in the range of $|c^{K_1} - 50| \leq 4$ when $n = 100$ and $|c^{K_1} - 50| \leq 3$ when $n = 120$. There is a large difference in the distribution where y_i is generated at $x_i < 50$ or $50 \leq x_i < 70$, but there is a small difference in the distribution where y_i is generated at $50 \leq x_i < 70$ or $70 \leq x_i$. Therefore, it is difficult to calculate the cutoff value of c^{K_2} close to the population cutoff value. When $n = 80$, the proportion of the cutoff value of c^{K_2} was greater than 50% in the range of $|c^{K_2} - 70| \leq 5$. When $n = 120$, the proportion of the cutoff value of c^{K_2} was greater than 60% in the range of $|c^{K_2} - 70| \leq 5$. It is considered that the cutoff value of c^{K_2} also has sufficiently usable performance. As a result, it was confirmed that the number of cutoff values was correctly judged and that the two cutoff values were calculated close to the population cutoff values.

4. COVID-19 DATA

We judged the number of cutoff values and calculated the corresponding cutoff value(s). We used the clinical outcomes data from 110 hospitalized COVID-19 patients treated with famotidine and cetirizine for at least 48 h [Hogan II et al. 2020], as presented in Table VII. This data are presented by Supplementary data of their paper. This study included patients who got dual medical treatment for the therapy of COVID-19 from April 3, 2020, to June 13, 2020. Famotidine and cetirizine were given 20 mg intravenously and 10 mg intravenously (or oral), respectively, at 12 h intervals. In this manuscript, the cutoff values were calculated using data from 93 patients, excluding 17 deaths. There were 13 patients with do-not-resuscitate (DNR), but the analysis set of this manuscript included DNR. Histamine-1 blocker (H1) antihistamines are for the management of allergies, and histamine-2 blockers (H2) receptor antagonists are for the treatment of gastrointestinal disorders. Recently, it

Table VI. Proportion of two cutoff values in range in Section 3.2

<i>n</i>	Pattern	c^{K_1}					c^{K_2}				
		49-51	48-52	47-53	46-54	45-55	69-71	68-72	67-73	64-76	65-75
80	10	65.7%	77.5%	83.8%	87.4%	89.8%	24.2%	34.2%	42.2%	48.8%	54.2%
	11	69.4%	81.8%	88.3%	92.3%	94.7%	26.8%	37.3%	45.6%	52.3%	58.0%
	12	66.8%	79.9%	86.9%	90.6%	92.9%	28.7%	39.6%	48.0%	54.6%	60.2%
	13	70.1%	81.4%	86.8%	89.8%	91.5%	25.4%	35.8%	43.9%	50.5%	55.9%
	14	73.8%	85.7%	91.3%	94.5%	96.3%	28.1%	39.1%	47.5%	54.3%	59.9%
	15	71.1%	83.9%	90.1%	93.3%	95.0%	29.9%	41.1%	49.6%	56.4%	62.0%
	16	79.9%	89.4%	93.6%	95.6%	96.7%	43.5%	56.9%	65.9%	72.2%	76.8%
	17	81.8%	91.5%	95.6%	97.7%	98.7%	47.1%	60.9%	69.6%	75.6%	80.1%
	18	79.4%	90.6%	95.4%	97.5%	98.3%	50.1%	63.5%	72.0%	77.7%	81.8%
100	10	72.4%	82.8%	87.7%	90.4%	92.2%	27.3%	37.8%	45.7%	52.1%	57.3%
	11	75.3%	86.0%	91.2%	94.0%	95.8%	30.1%	41.1%	49.4%	56.0%	61.3%
	12	75.3%	86.7%	92.2%	95.1%	96.6%	32.6%	44.1%	52.6%	59.2%	64.5%
	13	76.6%	86.1%	90.2%	92.3%	93.5%	28.8%	39.6%	47.7%	54.2%	59.4%
	14	79.9%	89.5%	93.6%	95.8%	97.0%	31.9%	43.3%	51.7%	58.3%	63.4%
	15	79.4%	90.0%	94.7%	96.8%	98.0%	34.3%	46.0%	54.5%	60.9%	66.2%
	16	85.9%	93.3%	96.1%	97.4%	98.1%	49.7%	63.0%	71.4%	77.1%	81.2%
	17	87.3%	94.5%	97.3%	98.6%	99.2%	53.7%	67.1%	75.2%	80.6%	84.4%
	18	86.9%	95.1%	98.0%	99.2%	99.6%	57.2%	70.4%	78.1%	83.1%	86.5%
120	10	77.7%	86.7%	90.7%	92.7%	94.0%	30.6%	41.2%	49.3%	55.7%	60.9%
	11	80.3%	89.4%	93.5%	95.6%	96.9%	33.8%	45.1%	53.4%	59.9%	65.1%
	12	80.0%	89.8%	94.4%	96.7%	97.9%	36.5%	48.3%	56.8%	63.3%	68.4%
	13	82.2%	89.9%	93.0%	94.5%	95.3%	32.8%	44.2%	52.5%	58.7%	63.8%
	14	84.7%	92.6%	95.7%	97.1%	97.9%	35.6%	47.5%	56.1%	62.5%	67.6%
	15	84.2%	92.9%	96.5%	98.1%	98.9%	38.7%	50.9%	59.5%	65.8%	70.7%
	16	90.2%	95.8%	97.8%	98.6%	99.0%	55.4%	68.6%	76.4%	81.6%	85.1%
	17	91.2%	96.7%	98.5%	99.3%	99.6%	59.4%	72.4%	79.9%	84.7%	88.0%
	18	90.7%	96.7%	98.7%	99.5%	99.8%	62.9%	75.4%	82.6%	86.9%	89.9%

Table VII. Clinical outcomes in 110 hospitalized COVID-19 patients (*x*: age (years old), *y*: days to discharge (day), ^a: days to death, ^b: patient with DNR).

<i>x</i>	79	53	34	64	78	50	83	71	85	91	73	65	81	57	93	79	71	59	50
<i>y</i>	5	6	2	32	18	5	11	4	5	33	35	14	18	8	12	8	9	4	5
<i>x</i>	43	80	58	39	46	41	60	68	89	83	39	72	45	63	87	43	92	22	92
<i>y</i>	7	20	29	7	8	6	7	11	16 ^{a,b}	14 ^{a,b}	18	16	15	11	6 ^{a,b}	7	12	10	11
<i>x</i>	64	72	92	72	51	81	56	74	64	58	57	70	17	38	81	69	51	51	80
<i>y</i>	10	21	6 ^b	5	11	20	5	6	8	6	13	7	7	10 ^{a,b}	6	42	9	11	4
<i>x</i>	61	80	25	63	89	76	24	71	69	97	27	71	76	66	60	79	84	63	49
<i>y</i>	25	11	10	11 ^{a,b}	5 ^{a,b}	5 ^{a,b}	7	10	19	6 ^{a,b}	6	9	5	9	4 ^{a,b}	7	7 ^b	6 ^a	6
<i>x</i>	94	79	68	63	69	91	79	61	48	33	76	50	37	21	53	73	56	67	45
<i>y</i>	17 ^b	5	30	13	20 ^a	14 ^a	10 ^a	12 ^a	7	15	19	4	3	4	12 ^a	13	8	5	11
<i>x</i>	73	75	73	43	55	68	63	48	38	70	60	73	57	75	72				
<i>y</i>	8	8	5	12	9 ^a	16	8	5 ^a	6	5	13	14 ^b	7	4	8				

has become clear that cetirizine, which contains the H1 component, and famotidine, which contains the H2 component, have been discovered to have a considerable anti-SARS-CoV-2 impact [Freedberg et al. 2020; Janowitz et al. 2020; Blanco et al. 2021].

Table VIII. Characteristics of three groups. Data are summarized as frequency (proportion).

	Age : <59 (<i>n</i> = 40)	Age : 60-69 (<i>n</i> = 22)	Age : 70- (<i>n</i> = 48)
Discharge	36(90.0%)	17(77.3%)	40(83.3%)
Death	4(10.0%)	5(22.7%)	8(16.7%)
Days to discharge ≤ 10	27(67.5%)	6(27.3%)	23(47.9%)

We use the software R to calculate the cutoff values by the WMW and KW tests, and a sample code is provided in Appendix. The independent variable x is the age (years), and the dependent variable y is the days to discharge (day). According to the minimum P -value approach, the one cutoff value by the WMW test is 60 years old, and the P -value using that one cutoff value is 0.011. The two cutoff values by the KW test are 60 and 70 years old, and the P -value using those two cutoff values is 0.004. As a result, the two cutoff values of 60 and 70 years old are adopted. Fig. 1 shows a scatter plot of the age and days to discharge, and the dashed line presents two cutoff values of 60 and 70 years old. Table VIII summarizes the characteristics of the three groups. In the group of less than 60 years old, 27 patients were discharged within 10 days, and a few patients were long days to discharge. Approximately 67.5% of all patients were discharged within 10 days. In the group of 60-69 years old, 6 patients were discharged within 10 days, and several patients were long days to discharge. Approximately 27.3% of all patients were discharged within 10 days. In the group of greater than or equal to 70 years old, 23 patients were discharged within 10 day, and some patients were long days to discharge. Approximately 47.9% of all patients were discharged within 10 days. In the proportion of death, the group of 60-69 years old was the highest proportion at 22.7%, the group of greater than or equal to 70 years old was the second highest proportion at 16.7%, and the group of less than 60 years old was the lowest proportion at 10.0%. As a result, it is reasonable to set the two cutoff values to 60 and 70 years old.

The numbers of the WMW and KW tests were performed 43 and 675 times, respectively. In the Bonferroni method, the significance level was divided by a total of 718 times for the two tests performed. For the commonly used significance level of 0.05 (two-sided), the P -value is compared to the adjusted significance level of $0.05/718 = 0.0000696$. Therefore, it was not significant.

5. CONCLUSIONS

The risk usually rises with age, but in clinical trials, the risk may appear to fall after a certain age. In this case, dividing the age into three groups by two cutoff values was

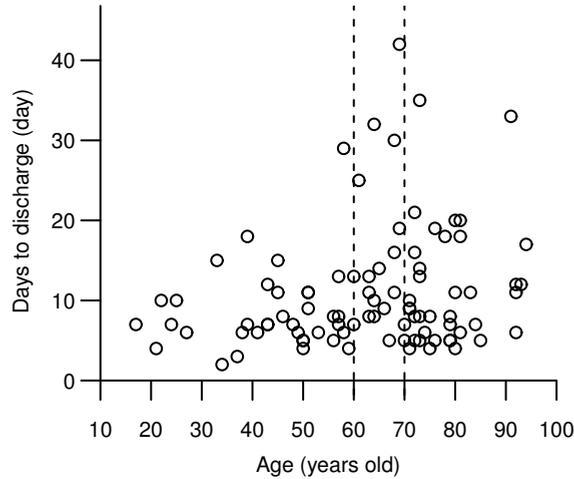


Fig. 1. Scatter plot of the age and days to discharge in COVID-19 data of 93 recovered patients. The dashed lines show two cutoff values.

thought to be more appropriate than dividing the age into two groups by one cutoff value. The one cutoff value was determined using the WMW test by the minimum P -value approach, and the two cutoff values were determined using the KW test by the minimum P -value approach. The smaller minimum P -value for the WMW and KW tests was adopted, and the cutoff value corresponding to that test was used. We validated the performance of this method using MCSs at various settings before applying it to COVID-19 data because it was a new method for determining the cutoff values. Using COVID-19 data, the minimum P -value by the WMW and KW tests were 0.011 and 0.004, respectively. The two cutoff values of 60 and 70 years old by the KW test were used. It was confirmed that each group had different characteristics by observing the data divided into three groups based on the two cutoff values. However, in the Bonferroni method, it was not significant. It was considered effective to determine the cutoff value(s) using the minimum P -value approach in the two types of tests, but the issue of reducing the number of tests remained. In this study, because we first calculated all the P -values at the potential cutoff value(s) and then searched for the smallest P -value, the number of tests performed was very large. It is conceivable that the number of tests performed can be reduced by improving the procedure. If the number of tests performed can be reduced, it may be significant even if the significance level is adjusted by the Bonferroni method. This is a future work.

ACKNOWLEDGMENTS

The authors are thankful learned reviewers for their valuable comments towards the improvements of the manuscript.

REFERENCES

- ALTMAN, D. G., LAUSEN, B., SAUERBREI, W., AND SCHUMACHER, M. 1994. Dangers of using “optimal” cutpoints in the evaluation of prognostic factor. *Journal of the National Cancer Institute* 86, 11, 829–835.
- BLANCO, J. I. M., BONILLA, J. A. A., HOMMA, S., SUZUKI, K., FREMONT-SMITH, P., AND DE LAS HERAS, K. V. G. 2021. Antihistamines and azithromycin as a treatment for COVID-19 on primary health care – a retrospective observational study in elderly patients. *Pulmonary pharmacology & therapeutics* 67, 101989.
- FREEDBERG, D. E., CONIGLIARO, J., WANG, T. C., TRACEY, K. J., CALLAHAN, M. V., ABRAMS, J. A., AND GROUP, F. R. 2020. Famotidine use is associated with improved clinical outcomes in hospitalized COVID-19 patients: a propensity score matched retrospective cohort study. *Gastroenterology* 159, 3, 1129–1131.
- GREINER, M., PFEIFFER, D., AND SMITH, R. D. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine* 45, 1–2, 23–41.
- HOGAN II, R. B., HOGAN III, R. B., CANNON, T., RAPPAL, M., STUDDARD, J., PAUL, D., AND DOOLEY, T. P. 2020. Dual-histamine receptor blockade with cetirizine-famotidine reduces pulmonary symptoms in COVID-19 patients. *Pulmonary pharmacology & therapeutics* 63, 101942.
- JANOWITZ, T., GABLENZ, E., PATTINSON, D., WANG, T. C., CONIGLIARO, J., TRACEY, K., AND TUVESON, D. 2020. Famotidine use and quantitative symptom tracking for COVID-19 in non-hospitalised patients: a case series. *Gut* 69, 9, 1592–1597.
- LIU, H., YANG, Y., CHEN, C., WANG, L., HUANG, Q., ZENG, J., LIN, K., ZENG, Y., GUO, P., ZHOU, W., AND LIU, J. 2020. Reclassification of tumor size for solitary HBV-related hepatocellular carcinoma by minimum p value method: a large retrospective study. *World journal of surgical oncology* 18, 1, 1–10.
- MAZUMDAR, M. AND GLASSMAN, J. R. 2000. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Statistics in medicine* 19, 1, 113–132.
- MIRKES, E. M., COATS, T. J., LEVESLEY, J., AND GORBAN, A. N. 2016. Handling missing data in large healthcare dataset: a case study of unknown trauma outcomes. *Computers in Biology and Medicine* 75, 203–216.
- NAKAMURA, T. 1986. BMDP program for piecewise linear regression. *Computer Methods and Programs in Biomedicine* 23, 1, 53–55.
- OGURA, T. AND SHIRAISHI, C. 2022. Cutoff value for Wilcoxon-Mann-Whitney test by minimum P -value: application to COVID-19 data. *International Journal of Statistics and Probability* 11, 3, 1–8.
- R CORE TEAM. 2021. R Foundation for Statistical Computing.
- REZNIKOV, L. R., NORRIS, M. H., VASHISHT, R., BLUHM, A. P., LI, D., LIAO, Y. S. J., BROWN, A., BUTTE, A. J., AND OSTROV, D. A. 2022. Identification of antiviral antihistamines for COVID-19 repurposing. *Biochemical and biophysical research communications* 538, 1, 173–79.
- ÜNÖZKAN, H., YILMAZ, M., AND DERE, A. M. 2020. A stochastic approach to number of corona virus cases. *Journal of Applied Mathematics, Statistics and Informatics* 16, 2, 67–83.
- VIETH, E. 1989. Fitting piecewise linear regression functions to biological responses. *Journal of applied physiology* 67, 1, 390–396.
- WORLD HEALTH ORGANIZATION. 2020. Coronavirus disease 2019 (COVID-19): situation report, 94.
- WORLD HEALTH ORGANIZATION. 2022. Coronavirus disease (COVID-19). <https://www.who.int/>.

ZOU, K. H., O'MALLEY, A. J., AND MAURI, L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115, 5, 654–657.

APPENDIX

We provided a sample code of software R that can be used to replicate the results. Another practical example can be performed by replacing two vectors of x and y with appropriate ones.

```
library(exactRankTests)
x<-c(79,53,34,64,78,50,83,71,85,91,73,65,81,57,93,79,71,59,50,43,80,
58,39,46,41,60,68,39,72,45,63,43,92,22,92,64,72,92,72,51,81,56,74,
64,58,57,70,17,81,69,51,51,80,61,80,25,24,71,69,27,71,76,66,79,84,
49,94,79,68,63,48,33,76,50,37,21,73,56,67,45,73,75,73,43,68,63,38,
70,60,73,57,75,72)
y<-c(5,6,2,32,18,5,11,4,5,33,35,14,18,8,12,8,9,4,5,7,20,29,7,8,6,7,
11,18,16,15,11,7,12,10,11,10,21,6,5,11,20,5,6,8,6,13,7,7,6,42,9,11,
4,25,11,10,7,10,19,6,9,5,9,7,7,6,17,5,30,13,7,15,19,4,3,4,13,8,5,
11,8,8,5,12,16,8,6,5,13,14,7,4,8)
n<-length(x); dat0<-data.frame(x,y); res1<-NULL; res2<-NULL
dat1<-dat0[order(dat0[,1]),]
for(c in (min(x)+1):max(x)){y1<-y[x<c]
y2<-y[x>=c]; if(length(y1)>=10&length(y2)>=10){res1<-rbind(res1,
c(c,wilcox.exact(y1,y2)$p.value))}}
res_1<-res1[order(res1[,2]),][1,]
for(c1 in (min(x)+1):(max(x)-1)){for(c2 in (c1+1):max(x)){y1<-y[x<c1
]; y2<-y[c1<=x&x<c2]; y3<-y[c2<=x]; if(length(y1)>=10&length(y2)>=
10&length(y3)>=10){g<-rep(NA,n); g[x<c1]<-1; g[c1<=x&x<c2]<-2; g[c2
<=x]<-3; res2<-rbind(res2,c(c1,c2,kruskal.test(y,g)$p.value))}}
res_2<-res2[order(res2[,3]),][1,]
res_all<-c(res_1,res_2,nrow(res1)+nrow(res2)); names(res_all)<-c("
Cutoff of WMW","Pvalue of WMW","Cutoff1 of KW","Cutoff2 of KW",
"Pvalue of KW","Number of tests performed")
res_all
```

Received August 2022; revised Month Year; accepted Month Year

Toru Ogura

Clinical Research Support Center

Mie University Hospital, 2-174, Edobashi, Tsu, 514-8507, Mie, Japan

Email: t-ogura@clin.medic.mie-u.ac.jp

Chihiro Shiraishi

Department of Pharmacy

Mie University Hospital, 2-174, Edobashi, Tsu, 514-8507, Mie, Japan

The application of PSO-BP combined model and GA-BP combined model in Chinese and V4's economic growth model

X. GUI, M. FEČKAN AND J. R. WANG

Abstract

This paper adopts different optimization algorithms such as Genetic Algorithm (GA) and Particle Swarm Optimization Algorithm (PSO-Algorithm) to train Back-Propagation (BP) neural networks, fits the Chinese, the Czech, Slovak, Hungarian, and Polish gross domestic product (GDP) growth model (from 1995 to 2020) and makes short-term simulation predictions. We use the PSO-Algorithm and GA with strong global search ability to optimize the weights and thresholds of the network, combine them with the BP neural network, and apply the resulting Particle Swarm Optimization Back-Propagation (PSO-BP) combined model or Genetic-Algorithm Back-Propagation (GA-BP) combined model to allow the network to achieve fast convergence. Besides, we also compare the above two hybrid models with standard multivariate regression model and BP neural network with different initialization methods like normal uniform and Xavier for fitting and short-term simulation predictions. Finally, we obtain the excellent results that all the above models have achieved a good fitting effect and PSO-BP combined model on the whole has a smaller error than others in predicting GDP values. Through the technology of PSO-BP and GA-BP, we have a clearer understanding of the five countries gross domestic product growth trends, which is conducive to the government to make reasonable decisions on the economic development.

Mathematics Subject Classification 2010: 26A33, 26A51, 26D15

General Terms: Algorithms

Keywords: Prediction, Gross domestic product, Genetic algorithm back-propagation combined model, Particle swarm optimization back-propagation combined model.

1. INTRODUCTION

The "16 + 1" model of cooperation between Central and Eastern European countries and China and the Belt and Road Initiative are in harmony in many aspects. Among central and eastern European countries, the Visegrad Four (V4), consisting of the Czech Republic, Hungary, Poland and Slovakia, has achieved great results in coordinating and strengthening national economies.

V4 countries are located in the heart of Europe, and trade and investment exchanges are increasing with the strengthening of cooperation and links between China and Central and Eastern Europe. At the same time, more and more V4 enterprises also look around the world, actively looking for new markets.

With the rapid development of global economy, import and export trade is

occupying an increasingly important position in the economy of various countries. The growth of national economy will promote the growth of import and export trade, and the growth of import and export trade will also play a role in promoting the development of national economy [1]. However, sometimes due to problems such as industrial structure and investment demand, the growth of import and export trade has little impetus for the development of the national economy. Therefore, we need to correctly select some nonlinear variables to study the GDP growth model to prevent policy makers from making inappropriate decisions.

In order to study V4's and China's economic growth model, this paper downloaded data sets related to above five countries GDP from 1995 to 2020 from the open database on the World Bank's official website. Due to the large dimensions of the initial data set, direct target research will consume too much time and resources, so this paper will select 7 explanatory variables according to [1; 2], and unify the measurement unit of the explanatory variables. Although GDP is a very complex nonlinear economic system, there are many good ways to study it, such as fractional calculus in [3; 4] which show that the fractional order model is superior to integer order model, and artificial neural network technology, especially the Back-Propagation (BP) neural network in [5; 6; 7; 8], which is widely used and mature. Recently, techniques such as artificial neural networks with flexible nonlinear modeling capabilities have gained popularity for dealing with nonlinearities in forecasting economic and financial time series [9; 10; 11]. In addition, there are many interesting models used to fit economic growth trends, such as grey multivariable forecasting models [12].

Although the BP neural network fits the model well, its limitations cannot be ignored. In order to solve the problem that it is easy to fall into the local extreme value and the network learning process is not stable, we have to propose two optimization algorithms with strong global optimization ability: Particle Swarm Optimization Algorithm (PSO Algorithm) [13; 14] and Genetic Algorithm (GA) [15; 16], which come from the research and observation of the life activities of bird predation and the law of biological evolution in nature, respectively. The global search feature of the PSO Algorithm and GA will make the BP neural network have a great prediction effect as [17; 18; 19; 20; 21; 22] have demonstrated. To the best of our knowledge, there are yet few results on forecasting the GDP by PSO-BP combined model and GA-BP combined model, which will be considered in this paper.

Then, we also compare the above two hybrid models with standard multivariate regression model and BP neural network with different initialization methods like normal uniform and Xavier. Though standard multivariate regression model and BP neural network with normal uniform initialization method (Uniform-*ini*-BP model) and BP neural network with Xavier initialization method (Xavier-*ini*-BP model) can get some good fitting and prediction results for GDP growth model, PSO-BP combined model and GA-BP combined model are better than three methods mentioned earlier. To be honest, owe to the lack of sufficient data which means it is difficult to come by older data of GDP and related variables, the prediction results of standard multivariate regression model, Uniform-*ini*-BP model and Xavier-*ini*-BP model are not good. Besides, PSO and GA are excellent search algorithms which can seek the best weights and thresholds in advance, though the small amount of data, the appropriate training structure of BP neural network still can be obtained.

Based on the comparison of mean square error (MSE), mean absolute deviation (MAD), the coefficient of determination (R^2), Bayesian information criterion (BIC) and images, it is concluded that the five methods are both suitable for this data set to fit the GDP growth model. Subsequently, the absolute relative error (ARE_i) will be used to evaluate the prediction results of the five models. In addition, the construction and training of all models, the calculation and operation process of GDP fitting and prediction in this paper will be implemented in MATLAB software.

The case study shows that the PSO-BP combined model on the whole has a smaller error than others in predicting GDP values.

Summarizing, this paper is arranged as follows. Section 2 briefly introduces some explanatory variables used for fitting GDP growth model, V4, PSO Algorithm, and GA. In Section 3, the results of fitting and prediction of GA-BP combined model and PSO-BP combined model for Chinese and V4's GDP values are shown. Finally, the conclusion is given in Section 4.

2. MODEL DESCRIPTION

We selected 7 explanatory variables according to [1; 2], which are different from the research of Ming et al. [4] that six variables were considered for their China's GDP Model. These explanatory variables are Land area (LA) (km²), Cultivated land (CL) (hectare), Total population (TP) (per), Export of goods and services (EGS) (2010 constant US dollars), Import of goods and services (IGS) (2010 constant US dollars), Final consumption expenditure of general government (FCEGG) (2010

constant US dollars), Gross capital formation (GCF) (2010 constant US dollars). The 26 data downloaded directly from the World Bank were from 1995 to 2020.

In order to simplify the expression, we define the following symbols:

Table I. Symbols.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	y	t
LA	CL	TP	EGS	IGS	FCEGG	GCF	GDP	year

2.1. PSO Algorithm And PSO-BP Combined Model

The basic idea of PSO is to design bird swarms as particles with only velocity and position properties. Velocity limits the speed of particle movement and position limits the direction of particle movement that Kennedy and Eberhart [13] first presented.

Each particle moves and searches within the specified search space, finds the optimal value, and then shares it with other individuals, and finally obtains the overall global optimal solution. Then we update the position and velocity of each particle according to the individual optimal solution of each particle and the global optimal solution obtained by all particles, and iterate until the upper limit of iteration is reached or the global optimal solution does not change for several iterations.

The basic step is that each particle first sets a random initial velocity and position as the initial optimal solution, then calculates the fitness value of each particle according to the set fitness, and selects the group optimal solution among all particles. If it does not reach the upper limit of the iteration or other constraints, then update the velocity and position of each particle, recalculate the fitness value of each particle, get the individual optimal solution, then update the global optimal solution, and iterate until the end condition is met.

As explicitly presented by Ethaib et al. [17], the velocity and position are updated by following formula:

$$V_{k+1} = \omega \times V_k + c_1 \times rand \times (pbest_k - x_k) + c_2 \times rand \times (gbest_k - x_k)$$

$$x_{k+1} = x_k + V_{k+1},$$

Where $k = 1, 2, 3, \dots, N$, N refers to the total number of particle swarms, ω is a non-negative inertia factor, V_k refers to the velocity of the k particle, and x_k is the position of the k particle. c_1 and c_2 are learning factors of individual information and group information respectively, generally take 2, $rand$ refers to a random number between 0

and 1, $pbest_k$ refers to the individual optimal solution searched by the k particle, and $gbest_k$ is the global optimal solution after all particles are searched.

Because the change of ω will have better results in the search for the optimal solution, the so-called linear weight will be used in this paper. The weight of each iteration is updated as follows:

$$\omega = (\omega_{mi} - \omega_{end}) \times \frac{k}{maxgen} + \omega_{end},$$

Where ω_{mi} is the initial inertia factor, ω_{end} is the maximum inertia factor, $maxgen$ is the maximum number of iterations, and k is the number of iterations.

The fitness function applied by this algorithm is shown below:

$$F = \sum_{i=1}^N |y_i - \hat{y}_i|,$$

Where N refers to the number of samples, y_i is the true value of the sample, and \hat{y}_i is the simulated prediction value of the network.

First, we give each particle a random position and velocity, and then use the fitness function given above to calculate the fitness value of the particle based on the particle position and the neural network layer structure which previously set. Then we select the individual optimal and the group optimal according to the fitness value of all particles. Then update the position and velocity of each particle with the position and velocity of the individual optimal and the group optimal, and calculate the position and velocity of each particle again and calculate their fitness value until the end of the iteration. It can be seen from the fitness function that the individual optimal and group optimal fitness values we are looking for should always decrease, and finally we get the final individual optimal and group optimal particles, and correspond their positions to thresholds and weights in the neural network, and then train the network.

The PSO-BP combined model in this paper firstly regards the weights and thresholds between nodes in each layer of the BP network as particles, and iterates the PSO algorithm to find the optimal solution, secondly converts to obtain the optimal weights and thresholds. Finally it constructs the BP network to learn from the data set and train a good network to make predictions.

2.2. Genetic Algorithm And GA-BP Combined Model

Genetic algorithm (GA) is an adaptive global optimization probabilistic search algorithm which simulates the genetic and evolutionary process of living organisms in the natural environment. There are many kinds of objective functions and

constraints in optimization problems, some are linear, some are nonlinear; Some are continuous, some are discrete; Some have single peaks, some have multiple peaks. With the deepening of the research, people gradually realize that it is impossible and unrealistic to find the optimal solution completely and precisely in many complicated cases, so it is one of the main focus of people to find the approximate optimal solution or satisfactory solution. Genetic algorithm provides an effective way and general framework to solve this kind of problem and creates a new global optimization search algorithm.

In genetic algorithm, n -dimensional decision vector $X = [x_1 x_2 \dots x_n]^T$ is represented by the symbol string X composed of n symbols X_i : $X = X_1 X_2 \dots X_n$. Considers each X_i as a genetic gene, and all of its possible values are called alleles. So X can be viewed as a chromosome made up of n genes. In general, the length of the chromosome n is fixed, but it can vary for some problems. Depending on the different cases, the allele can be a set of integers, a range of real values, or a pure marker. The simplest allele consists of two integers, 0 and 1, and the corresponding chromosome can be represented as a string of binary symbols. The arrangement X formed by this code is the genotype of the individual, and its corresponding X value is the phenotype of the individual. Usually an individual's phenotype and genotype correspond one to one, but sometimes a many-to-one relationship between genotype and phenotype is allowed. Chromosome X is also called individual X . For each individual X , its fitness should be determined according to certain rules. The individual fitness is correlated with the objective function value of the corresponding individual phenotype X . The closer X is to the optimum point of the objective function, the greater the fitness is; On the contrary, its fitness is smaller. The decision variable X in genetic algorithm constitutes the solution space of the problem. The search for the optimal solution of the problem is carried out through the search process of chromosome X , thus all chromosomes X constitute the search space of the problem.

The evolution of organisms is based on groups. The corresponding operation object of genetic algorithm is a set composed of M individuals, called population. Similar to the natural evolution of biology from generation to generation, the operation process of genetic algorithm is also an iterative process. The t generation population is denoted as $P(t)$, and the $t + 1$ generation population is obtained after one generation of heredity and evolution, which is also a set composed of multiple individuals and denoted as $P(t + 1)$. This population is continuously operated by heredity and evolution, and each time more individuals with higher fitness are inherited to the next generation according

to the rule of survival of the fittest. In this way, an excellent individual X will be finally obtained in the population, and its corresponding phenotype X will reach or close to the optimal solution X^* .

The evolutionary process of organisms is mainly accomplished by the crossover and mutation of chromosomes. Correspondingly, the search process for the optimal solution in the genetic algorithm also imitates the evolutionary process of organisms. The so-called genetic operator is applied to population $P(t)$ and the following genetic operations are performed to obtain the new generation population $P(t + 1)$.

1. Selection According to the fitness of each individual, some excellent individuals are selected from the t generation population $P(t)$ and inherited to the next generation population $P(t+1)$ according to certain rules or methods.
2. Crossover Each individual in a population $P(t)$ is paired randomly, and some chromosomes are exchanged parts of them between each pair at a probability (known as crossover rate).
3. Mutation For each individual in population $P(t)$, the gene value of one or some loci is changed to other alleles at a certain probability (called mutation rate).

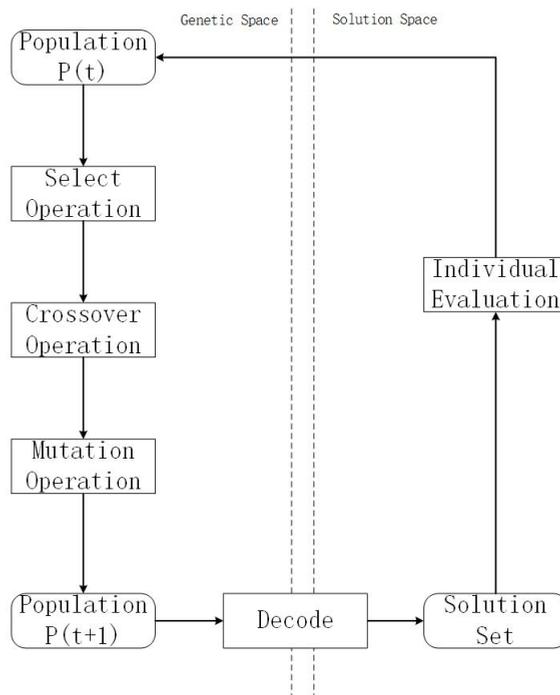


Fig. 1. The operation process of genetic algorithm.

The operation process of genetic algorithm is as follows (see Figure 1).

Step 1: Initialize. Set the evolution algebra counter $t = 0$, and the maximum evolution algebra T ; M individuals are randomly generated as the initial population $P(0)$.

Step 2: Individual evaluation. The fitness of each individual in population $P(t)$ is calculated.

Step 3: Select operation. The selection operator is applied to the population.

Step 4: Crossover operation. The crossover operator is applied to the population.

Step 5: Mutation operation. The mutation operator is applied to the population. The next generation population $P(t+1)$ is obtained after selection, crossover and mutation operation.

Step 6: Terminate condition judgment. If $t \leq T$, then $t = t + 1$, go to Step 2; If $t > T$, the individual with the maximum fitness obtained in the evolution process is taken as the optimal solution output and the calculation is terminated.

Similar to PSO-BP combined model, finally we can get the optimal solution output and an excellent individual X through operation process of genetic algorithm, which corresponds to thresholds and weights in the neural network. Subsequently, we can use them to train the network.

The GA-BP combined model encodes the weights and thresholds between nodes in each layer of the BP network as chromosomes, and iterates the GA algorithm to find the optimal solution X^* , and then decodes the chromosomes to obtain the optimal weights and thresholds. Finally it constructs the BP network to learn from the data set and train a good network to make predictions.

In order to compare the performance and predictive capabilities of the five models, in MATLAB software there is a toolbox for neural network modeling and related functions for training neural networks, which takes 70% of the data as the training set, 15% as the validation set, and 15% as the test set by default, i.e., 18 of the dataset as the training set, 4 of the dataset as the validation set, 4 of the dataset as the test set, respectively.

What needs to be added is that we have to normalize the data before putting the data into the research. We choose the minimum and maximum normalization to handle it, besides, generally we are fond of using the mean square error (MSE) to evaluate the fitting effect of the model, and the absolute relative error (ARE_i) to demonstrate the prediction effect between models.

Furthermore, we need the mean absolute deviation (MAD), the coefficient of

determination (R^2) and Bayesian information criterion (BIC) to explain further and evaluate the models that have good fitting effect. The formulas for minimum and maximum normalization, mean square error, absolute relative error, mean absolute deviation, the coefficient of determination and Bayesian information criterion are as follows:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}},$$

and

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

and

$$ARE_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right|, i = 1, 2, 3, \dots, n,$$

and

$$MAD = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n},$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2},$$

and

$$BIC = \log \left(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) + \frac{p \log n}{n}.$$

According to [23] and [24], The parameter settings of GA-BP combined model and PSO-BP combined model are as follows: (see Table II).

3. MAIN RESULTS

3.1. Model Evaluation

In this paper, by the reference parameter given in the [23] and [24] the network under GA and PSO algorithms is trained multiple times by continuously changing the parameters within reasonable range, and after many test experiments, we have obtained that the training network using the parameters of Table II is stable, which means that we can get the similar result within the reference parameter range. If the parameter setting exceeds range or is unreasonable, it will make the training network sensitive and make the fitting and prediction results have a large deviation. Following

Table II. Parameter setting.

Models	Parameters	Value
GA-BP	the individual number	40
	the binary number of bits of a variable	10
	the generation gap	0.95
	the maximum number of iterations	66
	the crossover rate	0.7
	the mutation rate	0.01
	three-layer structure	7-15-1
	hidden layer function	tansig
	output layer function	purelin
	learning rate	0.01
PSO-BP	learning factors c_1 and c_2	2.05
	the initial inertia factor ω_{ini}	0.9
	the maximum inertia factor ω_{end}	0.4
	the maximum number of iterations	200
	total number of particles	100
	maximum speed	1
	Minimum speed	-1
	maximum position	1
	minimum position	-1
	three-layer structure	7-3-1
hidden layer function	tansig	
output layer function	purelin	
learning rate	0.11	

the comparison of two algorithms convergence is shown in Fig. 2.

Through following figures, we can find that the convergence rate of GA is generally faster than that of PSO Algorithm because GA tends to converge prematurely. This leads to the possibility that convergence may not be optimal. In GA, chromosomes share information with each other, so the movement of the whole population is relatively uniform to the optimal region. Particles in PSO only share information through the current search to the optimum point, so to a large extent this is a single information sharing mechanism, and the whole search and update process is the process which follows the current optimal solution. In most cases, all particles may converge to the optimal solution more easily than evolutionary individuals in genetic algorithm.

3.2. Fitting Result

The PSO-BP combined model and GA-BP combined model above have the great effect on fitting the Chinese and V4's GDP model. Now we compare them with the standard multivariate regression model, Uniform-ini-BP model and Xavier-ini-BP model for the fitting effect. Through the training data set (original data), we

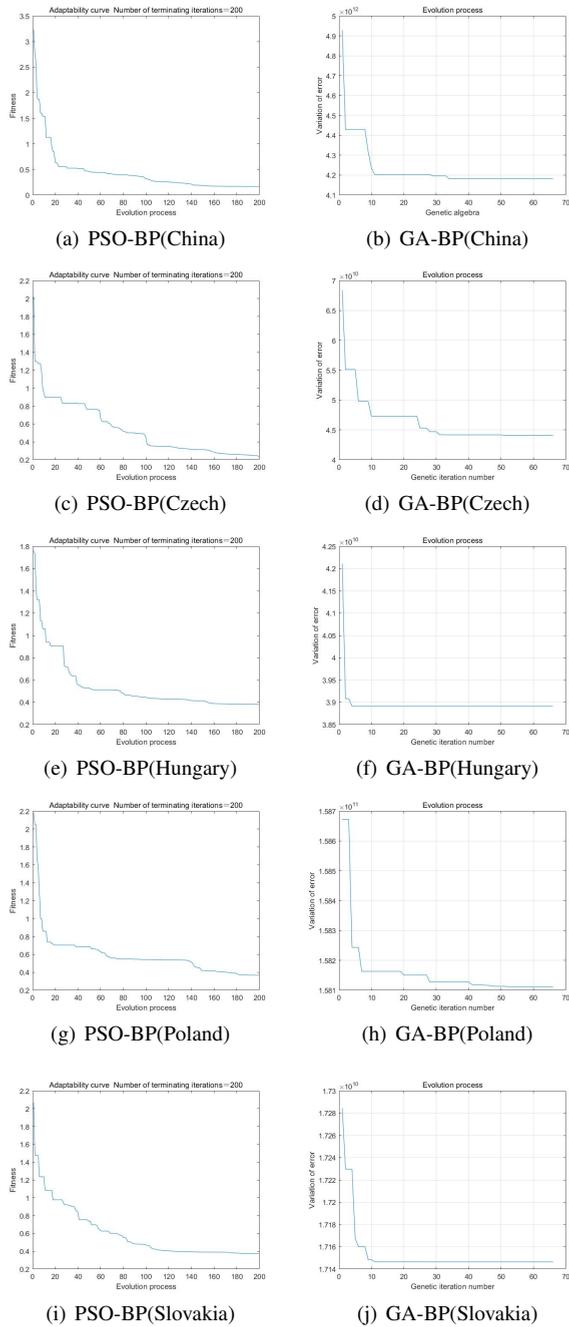


Fig. 2. GA and PSO-Algorithm convergence.

calculated the values of MSE, MAD, R^2 , and the BIC index to evaluate fitting effects of five models (see Table III).

Table III. Fitting performance of different models.

	R^2	MSE	MAD	BIC
GA-BP(China)	0.9986	7.4417×10^{21}	6.7997×10^{10}	51.3762
PSO-BP(China)	0.9989	1.2098×10^{22}	6.2015×10^{10}	51.7245
Xavier-BP(China)	0.9991	4.9506×10^{21}	5.3216×10^{10}	50.9686
Uniform-BP(China)	0.9955	2.4462×10^{22}	1.1378×10^{11}	52.5662
Regress-BP(China)	0.9998	8.6166×10^{20}	2.3910×10^{10}	49.2202
GA-BP(Czech)	0.9993	6.2455×10^{17}	6.0779×10^8	41.9907
PSO-BP(Czech)	0.9992	1.0431×10^{18}	6.1678×10^8	42.3659
Xavier-BP(Czech)	0.9945	4.7933×10^{18}	1.5297×10^9	44.0286
Uniform-BP(Czech)	0.9978	1.9381×10^{18}	1.0766×10^9	43.1231
Regress-BP(Czech)	0.9982	1.5339×10^{18}	9.7646×10^8	42.8892
GA-BP(Hungary)	0.9968	9.5391×10^{17}	8.2091×10^8	42.4142
PSO-BP(Hungary)	0.9991	4.5041×10^{17}	4.6647×10^8	41.5261
Xavier-BP(Hungary)	0.9921	2.3240×10^{18}	1.0905×10^9	43.3047
Uniform-BP(Hungary)	0.9943	1.7001×10^{18}	1.0027×10^9	42.9920
Regress-BP(Hungary)	0.9940	1.7622×10^{18}	9.2264×10^8	43.0280
GA-BP(Poland)	0.9991	7.4409×10^{18}	1.8016×10^9	44.4684
PSO-BP(Poland)	0.9998	3.6821×10^{18}	1.5408×10^9	43.6272
Xavier-BP(Poland)	0.9971	2.4656×10^{19}	2.5454×10^9	45.6664
Uniform-BP(Poland)	0.9947	4.4838×10^{19}	4.8814×10^9	46.2644
Regress-BP(Poland)	0.9962	3.2495×10^{19}	4.9007×10^9	45.9424
GA-BP(Slovakia)	0.9930	2.3125×10^{18}	1.2514×10^9	43.2997
PSO-BP(Slovakia)	0.9983	8.1206×10^{17}	4.9813×10^8	42.1155
Xavier-BP(Slovakia)	0.9951	1.6016×10^{18}	8.5852×10^8	42.9324
Uniform-BP(Slovakia)	0.9916	2.7586×10^{18}	1.0249×10^9	43.4761
Regress-BP(Slovakia)	0.9927	2.4015×10^{18}	1.3094×10^9	43.3375

As can be seen from the above table, the fitting results of five models for GDP models of five countries are so excellent. But the fitting effect of PSO-BP combined model in general is steadily better than other models. Now, we give the fitting and prediction results of the five models for the Chinese and V4's GDP growth model on MATLAB software (see Fig. 3, 4, 5, 6 and 7).

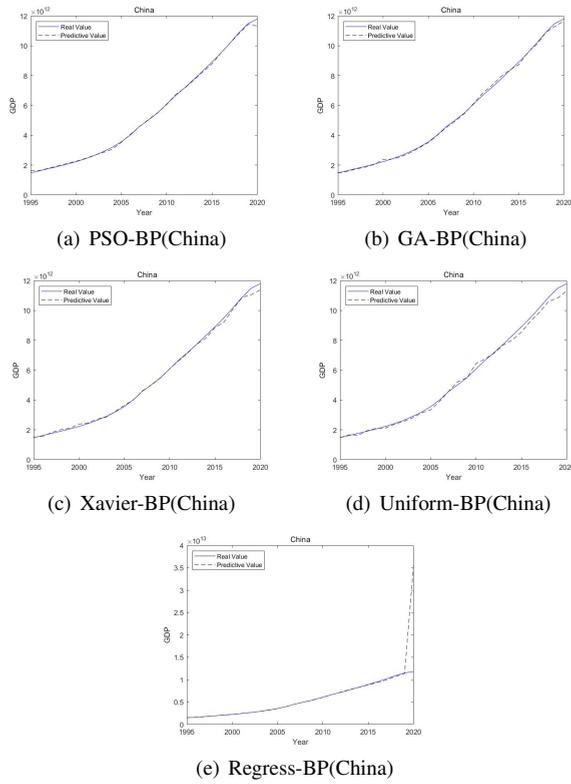


Fig. 3. Fitting and prediction results of the five models for the Chinese GDP growth model.

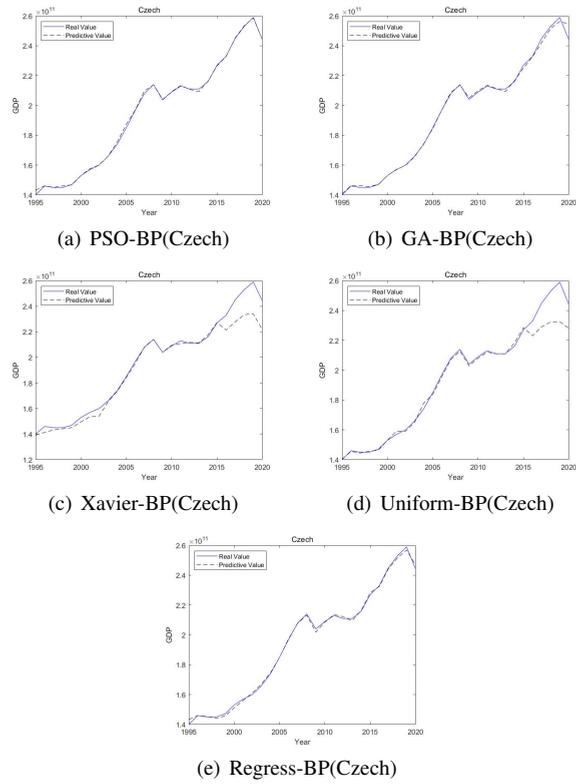


Fig. 4. Fitting and prediction results of the five models for the Czech GDP growth model.

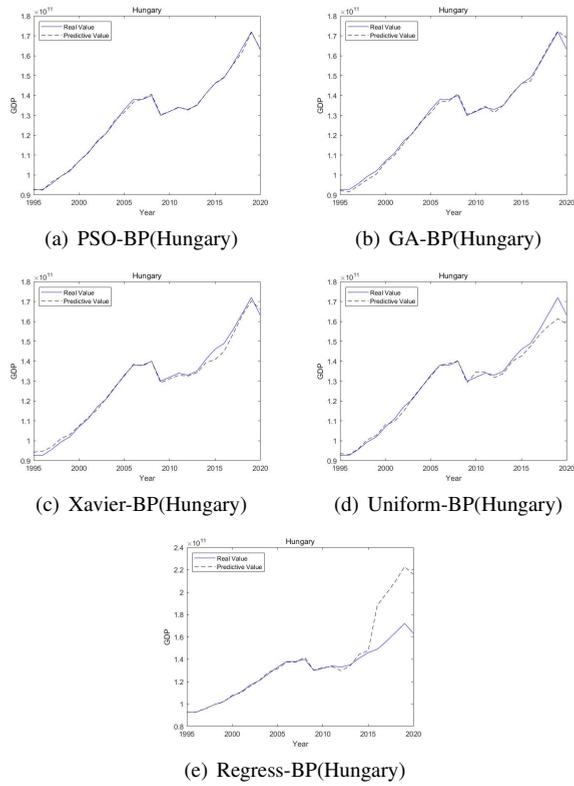


Fig. 5. Fitting and prediction results of the five models for the Hungarian GDP growth model.

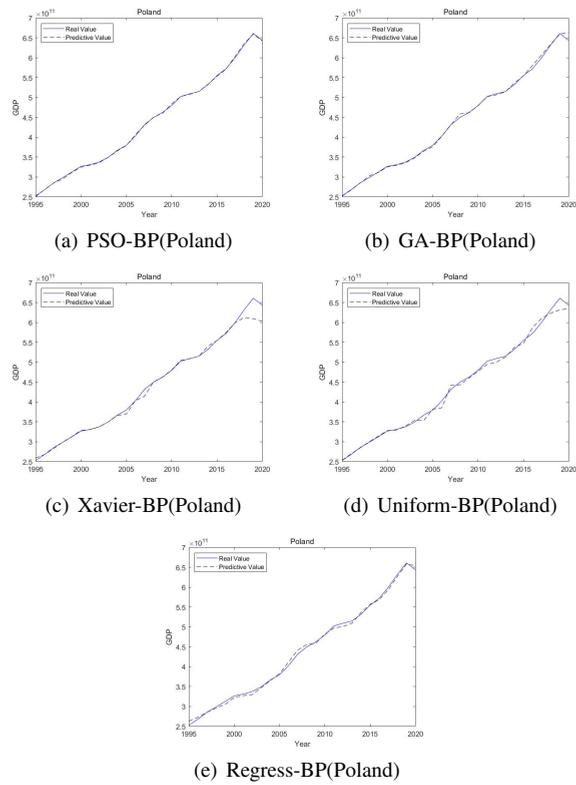


Fig. 6. Fitting and prediction results of the five models for the Polish GDP growth model.

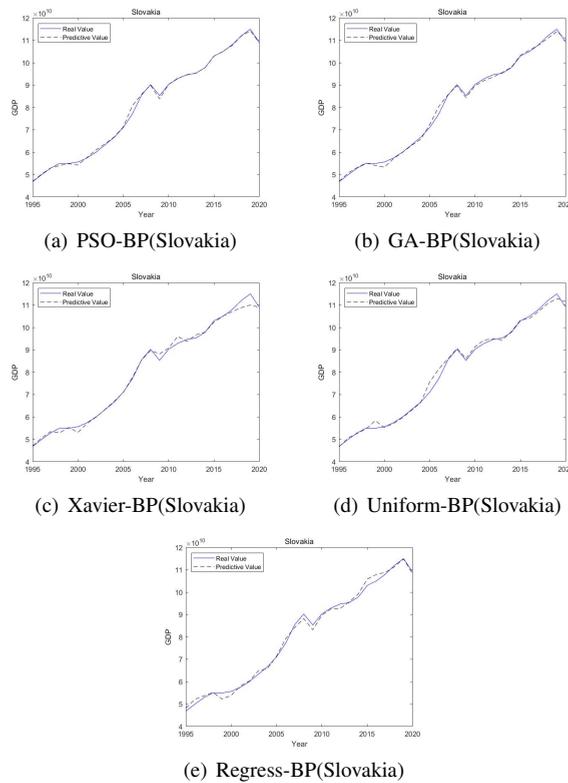


Fig. 7. Fitting and prediction results of the five models for the Slovakia’s GDP growth model.

It can be seen that the fitting results of the above graphs are great. And we can see from the graphs that there was a sharp drop between 2008 and 2009, which is resulting from the onset of the global financial tsunami in 2008. Subsequently, the Czech, Slovakian and Hungarian economies weakened and fell into recession. However, the GDP of Poland and China has not been affected. This is because their government has strengthened macro-control and introduced a lot of policies and measures to maintain the stable development of the financial market and the national economy, such as stock market policies, real estate policies and policies to expand domestic demand.

3.3. Predicted Result

In addition, from Fig. 3, 4, 5, 6 and 7, it is easier to understand that the prediction results of PSO-BP combined model and GA-BP combined model are better than the other three models. Actually, in the standard multivariate regression model occur steep rises and drops and loss the ability of excellent prediction due to the lack of sufficient

data and miscalculated regression coefficients for Chinese and Hungarian GDP growth models. Similarly, Uniform-*ini*-BP model and Xavier-*ini*-BP model produce poor prediction for Czech GDP growth model because of the lack of sufficient data. In fact, the data are nearly 30 years of data obtained from the World Bank, and it is difficult to come by older data of GDP and related variables. In addition, for these three models, at least hundreds of data are needed to make the model fit and predict effectively.

Finally, we present the forecast results of the PSO-BP combined model, GA-BP combined model, the standard multivariate regression model, Uniform-*ini*-BP model and Xavier-*ini*-BP model for Chinese and V4's GDP data from 2016-2020, and we calculate the ARE_i index values, as shown in Table IV, V and VI.

Table IV. Predicted Results

Country	Year	Real Value	GA-BP		PSO-BP	
			Predicted value	ARE_i	Predicted value	ARE_i
China	2016	952000000000	9559829702636.29	0.4184%	9525836035575.34	0.0613%
	2017	10200000000000	10095351712021.5	1.0260%	10194955074611.1	0.0495%
	2018	10900000000000	10898918408788.7	0.0099%	10837494585059.9	0.5734%
	2019	11500000000000	11304121203798.9	1.7033%	11436472069898	0.5524%
	2020	11800000000000	11667830275208	1.1201%	11323460534175.4	4.0385%
Czech	2016	233000000000	232671067729.118	0.1412%	232799057812.847	0.0862%
	2017	245000000000	242166639211.853	1.1565%	245237693138.668	0.0970%
	2018	253000000000	251630590796.729	0.5413%	253372605707.976	0.1473%
	2019	259000000000	256356783233.144	1.0205%	258673182906.291	0.1262%
	2020	244000000000	254528108385.872	4.3148%	244037849551.354	0.0155%
Hungary	2016	149000000000	147222609308.264	1.1929%	149225865393.829	0.1516%
	2017	156000000000	156379939797.03	0.2436%	155691936900.501	0.1975%
	2018	164000000000	164879828261.702	0.5365%	161931174141.02	1.2615%
	2019	172000000000	172327716942.825	0.1905%	171715393538.019	0.1655%
	2020	163000000000	168780508818.201	3.5463%	163377263177.46	0.2314%
Poland	2016	572452000000	580840498210.459	1.4654%	571508221331.321	0.1649%
	2017	600105000000	607134900727.096	1.1714%	603610194683.615	0.5841%
	2018	632233000000	633877597925.932	0.2601%	635918197790.091	0.5829%
	2019	660942000000	659467406246.733	0.2231%	659634332214.273	0.1978%
	2020	643085000000	663959185040.113	3.2459%	641970713012.838	0.1733%
Slovakia	2016	105000000000	105627733758.945	0.5978%	105106171819.703	0.1011%
	2017	108000000000	107875777395.503	0.1150%	107737968422.945	0.2426%
	2018	112000000000	110935156198.492	0.9508%	111926769158.625	0.0654%
	2019	115000000000	113922528919.196	0.9369%	114220595181.397	0.6777%
	2020	109000000000	110549200606.95	1.4213%	108833063528.208	0.1532%

Table V. Predicted Results

Country	Year	Real Value	Xavier-BP		Uniform-BP	
			Predicted value	ARE_i	Predicted value	ARE_i
China	2016	9520000000000	9228764653862.61	2.5080%	9281239832597.31	3.0592%
	2017	10200000000000	10108892724016.8	0.8932%	9897834397247.79	2.9624%
	2018	10900000000000	10881335424721.6	0.1712%	10617829278390.6	2.5887%
	2019	11500000000000	11062097074100.7	3.8079%	10840373967059.3	5.7359%
	2020	11800000000000	11385613714636.8	3.5117%	11325788465426.8	4.0187%
Czech	2016	2330000000000	221393149620.95	4.9815%	223053956790.816	4.2687%
	2017	2450000000000	227780141457.145	7.0285%	229081112144.987	6.4975%
	2018	2530000000000	233367425365.563	7.7599%	232292290367.562	8.1849%
	2019	2590000000000	234049553706.394	9.6334%	232453855081.095	10.2495%
	2020	2440000000000	221537241566.541	9.2060%	228016540899.785	6.5506%
Hungary	2016	1490000000000	145016273023.257	2.6736%	147576671026.596	0.9553%
	2017	1560000000000	153836901950.414	1.3866%	153348823467.122	1.6995%
	2018	1640000000000	163173867448.17	1.3866%	157826955376.149	1.6995%
	2019	1720000000000	170265939644.458	1.0082%	161375535719.657	6.1770%
	2020	1630000000000	166067725708.437	1.8820%	158728372248.949	2.6206%
Poland	2016	572452000000	574894936428.818	0.4267%	587185088957.002	2.5737%
	2017	600105000000	599594243608.192	0.0851%	610468620631.883	1.7270%
	2018	632233000000	612419294480.916	3.1339%	623234533416.615	1.4233%
	2019	660942000000	609387992750.376	7.8001%	631723348369.546	4.4208%
	2020	643085000000	602528603708.154	6.3065%	634416295008.516	1.3480%
Slovakia	2016	105000000000	105095878231.988	0.0913%	104027851181.114	0.9259%
	2017	108000000000	107063587087.942	0.8670%	107178057715.004	0.7611%
	2018	112000000000	108855433260.537	2.8076%	110656797265.638	1.1993%
	2019	115000000000	109996218558.591	4.3511%	112915358213.28	1.8127%
	2020	109000000000	108951741296.439	0.0443%	111480581571.937	2.2758%

Table VI. Predicted Results

Country	Year	Real Value	Regress-BP	
			Predicted value	ARE_i
China	2016	952000000000	9362829800941.65	1.6509%
	2017	1020000000000	9953806835333.37	2.4137%
	2018	1090000000000	10695427646463.8	1.8768%
	2019	1150000000000	11370943300657	1.1222%
	2020	1180000000000	35740405395443.5	202.8848%
Czech	2016	233000000000	232489823588.05	0.2190%
	2017	245000000000	244393091121.295	0.2477%
	2018	253000000000	251606001155.988	0.5510%
	2019	259000000000	256889253975.867	0.8150%
	2020	244000000000	246883607497.955	1.1818%
Hungary	2016	149000000000	188297943815.426	26.3745%
	2017	156000000000	198809035052.238	27.4417%
	2018	164000000000	209706427695.974	27.4417%
	2019	172000000000	222799336865.215	29.5345%
	2020	163000000000	215577057933.806	32.2559%
Poland	2016	572452000000	568561487876.142	0.6796%
	2017	600105000000	594215632169.578	0.9814%
	2018	632233000000	627040232007.714	0.8213%
	2019	660942000000	658623753950.821	0.3507%
	2020	643085000000	653485270607.433	1.6172%
Slovakia	2016	105000000000	107918413360.082	2.7794%
	2017	108000000000	108827715124.62	0.7664%
	2018	112000000000	111351392129.538	0.5791%
	2019	115000000000	114813824661.321	0.1619%
	2020	109000000000	108107379513.637	0.8189%

After a large number of repeated experiments on the same parameters in Table II, although sometimes the prediction effects of the PSO-BP combined model and GA-BP combined model are close (as shown in the prediction results for 2016-2020 in the Table IV, V and VI), the PSO-BP combined model requires a lot of repeated training times than the GA-BP combined model to obtain good results as shown in Fig. 2.

Further, the results in Table IV, V and VI demonstrate the prediction effect of Uniform-ini-BP model and Xavier-ini-BP model is poor again. However, the standard multivariate regression model has a good prediction for Slovakian, Polish and Czech GDP growth model.

In addition, the PSO-BP combined model and GA-BP combined model both have good forecasting effects, but there is a phenomenon that both of them are only suitable for short-term forecasting and will have large deviations in the long-term forecasting, such as in 2020 the PSO-BP combined model forecasts China's GDP, and GA-BP combined model forecasts the GDP of the Czech Republic, Hungary and Poland. In

general, we conclude that the prediction effect of PSO-BP combined model is better than that of GA-BP combined model.

Besides, we can see from Table IV that the PSO-BP mixed model has a large error in the forecast value of China's GDP in 2020, because the COVID-19 outbreak in Wuhan, China at the end of 2019 led to a very low economic development of the whole China in 2020. In addition, the standard multivariate regression model in Table VI has the same reason for the large prediction error of China's GDP in 2020. The COVID-19 pandemic leads to the small benefit of some GDP related variables with large regression coefficients in 2020, which makes the predicted value much higher than the actual value.

The GDP values of the V4 countries in Tables IV, V and VI for 2020 also fluctuated, as economic recession and mass bankruptcies became more real, and inflation and the energy crisis strained the economic situation. We can also see that in recent years, due to the initiative of the Belt and Road, V4 countries actively change their industrial structure, vigorously develop industry and service industry, and get a lot of foreign investment at the same time, the opportunity for economic exchanges with countries around the world also increased a lot.

4. CONCLUSION

We trained the BP neural network based on PSO and GA algorithm respectively with the data set composed of 7 explanatory variables, and then fitted Chinese and V4's GDP growth from 1995 to 2020. The graph of algorithm convergence shows that GA will converge ahead of PSO-Algorithm to find the global optimal solution, and bring the optimal threshold and weight value into the network to obtain the best BP neural network. Although the GA-BP model has the phenomenon of premature convergence, the optimal solution can be found and the prediction effect is excellent for the problems that are not too complicated. Compared with GA, PSO-Algorithm does not require coding, crossover and mutation operations, and particles are just updated by internal speed. Therefore, the PSO-Algorithm is simpler, parameters are fewer, and easier to implement.

Then we consider making a comparison about the above two algorithms and the standard multivariate regression model, normal uniform initialization method and Xavier initialization method for the fitting effect. It can manifest from the fitting graph that the five models are very effective. However, according to the prediction results of Chinese and V4's GDP from 2016 to 2020, and compared with their actual

values, it reveals that the absolute relative error of the PSO-BP combined model is smaller, and its prediction effect is better, while the prediction results of Uniform-ini-BP model, Xavier-ini-BP model and multivariate regression model are poor.

Then the best model we get can be used to predict future economic development, so that government decision-makers can make appropriate decisions, which will benefit the economic development of our country. However, for a global public emergency such as the COVID-19 pandemic in 2020, which has a great impact on the economy, the variables we selected are likely to be insufficient, which will lead to large errors in the trained models [25]. Therefore, in the following studies, we should take into account the influence factors of the novel coronavirus epidemic and further update the models to make the economic forecast of the next few years more accurate. This will help the government to introduce corresponding policies to optimize the country's industrial structure and solve the problems of inflation and resource shortage, which lead to rapid economic recovery.

REFERENCES

- R. E. Lucas, On the mechanics of economic development, *Journal of Monetary Economics*, 22 (1988), 3C42.
- W.J. Baumol, et al, *Convergence of Productivity*, Oxford University Press, 1994, 20C61.
- D. Luo, J. Wang, M. Fečkan, Applying fractional calculus to analyze economic growth modelling, *Mathematics*, 14(2018), 25-36.
- H. Ming, J. Wang, M. Fečkan, The application of fractional calculus in Chinese economic growth models, *Mathematics*, 7(2019), 665.
- M. Ilie, N. Popovici, C. Ilie, Simulation with artificial intelligence to forecast gdp depending on logistics elements, *International Management Conference, Faculty of Management, Academy of Economic Studies, Bucharest, Romania*, 2015.
- Z. Xiao, S. Ye, B. Zhong, et al, BP neural network with rough set for short term load forecasting, *Applied Artificial Intelligence Review*, 36(2009), 273-279.
- L. Feng, J. Zhang, Application of artificial neural networks in tendency forecasting of economic growth, *American Economic Journal Economic Policy*, 40(2014),76-80.
- X. Wang, J. Wang, M. Fečkan, BP neural network calculus in economic growth modelling of the group of seven, *Mathematics*, 8(2020), 11.
- C. Chuku, A. Simpasa, J. Oduor, Intelligent forecasting of economic growth for developing economies, *International Economics*, 159(2019), 74-93.
- S. Mladenović, M. Milovančević, I. Mladenović, et al, Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters, *Computers in Human Behavior*, 65(2016), 43-45.

-
- L. Feng, J. Zhang, Application of artificial neural networks in tendency forecasting of economic growth, *Economic Modelling*, 40(2014), 76-80.
- C. Wang, Y. Cao, Forecasting Chinese economic growth, energy consumption, and urbanization using two novel grey multivariable forecasting models, *Journal of Cleaner Production*, 299(2021), 126863.
- R. Eberhart, J. Kennedy. A new optimizer using particle swarm theory, MHS'95, Proceedings of the Sixth International Symposium on Micro Machine and Human Science, IEEE, 2002.
- K. Y. Lee, M. A. El-Sharkawi, Modern heuristic optimization techniques, *Fundamentals of Particle Swarm Optimization Techniques*, 2008, 71-87.
- M. Alba, R. Fernando, Evaluating genetic algorithms through the approximability hierarchy, *Journal of Computational Science*, 2021, 53, 101388.
- J. Holland, *Adaptation in natural and artificial systems : an introductory analysis with application to biology*, Control & Artificial Intelligence, 1975.
- S. Ethaib, R. Omar, M. K. S. Mazlina, et al, Development of a hybrid PSOCANN model for estimating glucose and xylose yields for microwave-assisted pretreatment and the enzymatic hydrolysis of lignocellulosic biomass, *NEURAL COMPUTING & APPLICATIONS*, 2016.
- Y. Mei, J. Yang, Y. Lu, et al, BP-ANN model coupled with particle swarm optimization for the efficient prediction of 2-chlorophenol removal in an electro-oxidation system, *International Journal of Environmental Research and Public Health*, 16(2019), 2454.
- F. van-den-Bergh, A. P. Engelbrecht, Using cooperative particle swarm optimization to train product unit neural networks, *IEEE International Joint Conference on Neural Networks*, Washington D C, USA, 2001.
- E. Assareh, M. A. Behrang, M. R. Assari, et al, Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran, *Energy*, 35(2010), 5223-5229.
- K. Harun, C. Halim, H. Arif, et al, Estimating petroleum exergy production and consumption using vehicle ownership and GDP based on genetic algorithm approach, *Renewable and Sustainable Energy Reviews*, 8(2004), 289-302.
- J. Nishtha, S. Bharti, Particle Swarm and Genetic Algorithm applied to mutation testing for test data generation: A comparative evaluation, *Journal of King Saud University - Computer and Information Sciences*, 32(2020), 514-521.
- X. Wang, *Introduction to neural networks*, Science Press, Beijing, China, 12(2016), 1-108. (Chinese)
- I. C. Trelea, The particle swarm optimization algorithm: convergence analysis and parameter selection, *Information Processing Letters*, 85(2003), 317-325.
- Y. Lyu, J. Nie, S. X. Yang, Forecasting US economic growth in downturns using cross-country data, *Economics Letters*, 198(2021), 109668.

Xudong Gui
4 Department of Mathematics
Guizhou University
Guiyang, Guizhou 550025, P.R. China
e-mail: xdguimath@126.com

Michal Fečkan
Department of Mathematical Analysis and Numerical Mathematics
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava,
Slovakia
and
Mathematical Institute of Slovak Academy of Sciences
Štefánikova 49, 814 73 Bratislava, Slovakia
e-mail: Michal.Feckan@fmph.uniba.sk

JinRong Wang
Department of Mathematics
Guizhou University
Guiyang, Guizhou 550025, P.R. China
e-mail: wjr9668@126.com; sci.jrwang@gzu.edu.cn

A heteroscedastic Bayesian model for method comparison data

S. M. M. LAKMALI, L. S. NAWARATHNA AND P. WIJEKOON

Abstract

When implementing newly proposed methods on measurements taken from a human body in clinical trials, the researchers carefully consider whether the measurements have the maximum accuracy. Further, they verified the validity of the new method before being implemented in society. Method comparison evaluates the agreement between two continuous variables to determine whether those measurements agree on enough to interchange the methods. Special consideration of our work is a variation of the measurements with the magnitude of the measurement. We propose a method to evaluate the agreement of two methods when those are heteroscedastic using Bayesian inference since this method offers a more accurate, flexible, clear, and direct inference model using all available information. A simulation study was carried out to verify the characteristics and accuracy of the proposed model using different settings with different sample sizes. A gold particle dataset was analyzed to examine the practical viewpoint of the proposed model. This study shows that the coverage probabilities of all parameters are greater than 0.95. Moreover, all parameters have relatively low error values, and the simulation study implies the proposed model deals with the higher heteroscedasticity data with higher accuracy than others. In each setting, the model performs best when the sample size is 500.

General Terms: Method Comparison, Simulation, Bayesian Inference.

Mathematics Subject Classification 2010: 62Exx, 62Hxx

Keywords: Agreement, Bayesian modelling, Concordance Correlation, Heteroscedastic Measurements, Total deviation index.

1. INTRODUCTION

The method comparison studies are performed to validate each newly introduced method before it is used in a clinical trial (Altman and Bland 1983; Andrew et al 2004; Aravind and Nawarathana 2017). Therefore, method comparison studies are vital for health-related fields like medicine, medical imaging, and biomedical engineering (Bland and Altman 1999; Boscardin and Gelman 1994). Also, critical consideration should be taken of clinical measurements related to the human body like blood pressure, heart rate, cholesterol level, etc. Obtaining correct measurements related to the human body is important, since humans' lives directly affect these measurements (Bilić-Zulle 2011). Although with the rapid development of technology, new methods

¹ The authors have no competing interests to declare that are relevant to the content of this article, and no funding was received to assist with the preparation of this manuscript.

can be introduced under several conditions and assumptions. Therefore, the agreement of existing and newly introduced methods should be checked, and the users prefer to use cheaper, less invasive, and user-friendly methods in general (Altman and Bland 1986, 1999). These methods may be based on a medical device, a clinical observer, or an instrument used to measure the measurements (Altman and Bland 1999; Boscardin and Gelman 1996). Before making decisions to use the methods interchangeably, it is necessary to conduct an accurate analysis to check the agreement of measurements. Method comparison evaluates the agreement between two continuous variables taken from two different methods, whether those measurements agree well enough to interchange the methods (Boscardin and Gelman 1994, 1996; Carstensen 2010).

Method comparison has developed under different conditions with different theories to answer the rising problems in agreement evaluation (Bilić-Zulle 2011; Altman and Bland 1986; Boscardin and Gelman 1996; Chinchilli et al 1996; Choudhary and Nagaraja 2007). It is started by calculating the difference of measurements from two methods in the standard subject using the bias. Limit of Agreement (LOA) is one of the most commonly used measures in method comparison studies. The perfect agreement can measure by the measurement pair with 450 lines through the origin when the difference of measurements is zero (Boscardin and Gelman 1996; Carstensen 2010; Chinchilli et al 1996). Although graphical illustration gives visualization results for method comparison, measures like Total Deviation Index (TDI), Concordance Correlation Coefficient (CCC) (Lin 1989; Lawrence et al 2007), Intra-class Correlation Coefficient (ICC) (Chinchilli et al 1996; Choudhary and Nagaraja 2007; Choudhary 2009; Choudhary and Yin 2010) intuitively appealing ideas about the agreement of measurements (Carstensen 2010). Moreover, various regression models like Deming regression (Altman and Bland 1983; Bilić-Zulle 2011) and Passing Bablok regression (Altman and Bland 1983; Bland and Altman 1986) also involve method comparison.

The homoscedastic measurements have many models for measuring the agreement, which has constant error variance. But this paper considered the heteroscedastic measurements, which have error variability changes with the

magnitude of the measurement. Heteroscedasticity is the Greek word used to mean the violation of homoscedasticity (Studenmund 1991). It is always discussed in the context of the residuals or error term. Specifically, heteroscedasticity is a systematic change in the spread of the residuals over the range of measured values. Scatter plots are the easiest way of identifying heteroscedasticity using the visual method because heteroscedastic data tend to follow a cone shape on a scatter plot of the residuals. Further, there are some statistical tests to confirm heteroscedasticity. Non-constant error variance test (NCV) and Breush-Pagan test are the most common tests for detecting heteroscedasticity of the variable. In most practical situations, heteroscedasticity becomes the big issue of modelling data. Although transformation helps to solve this case, it causes to create more complicated problems, especially in interpretation (Boscardin and Gelman 1994; Galea-Rojas et al. 2003).

Several models in method comparison deal with heteroscedastic measurements when going through the past literature. A model is proposed in (Galea-Rojas et al. 2003), using a heteroscedasticity measurement error model for method comparison only for replicate data. Although the heteroscedasticity measurement error model normally compares two methods, this can be used for more than two models. The model proposed in (Galea-Rojas et al. 2003) also can deal with the skewed and heavy-tailed ness by replacing the normality assumption with the generalization of normal. As the assumption of this model, an equal variance of both two methods is considered. This assumption can be considered as the limitation of this work. Copula-based methodology (Bland and Altman 1999) was also introduced for method comparison when dealing with heteroscedastic measurements. In (Bland and Altman 1999) the multivariate probability distribution is used, and the marginal probability distribution of each variable is the uniform distribution. This method is not assuming normality and constant error variance, and it is limited to non-replicated measurements. The major drawback of existing models is that these models can't deal with the non-replicated data. And also, the paper (Galea-Rojas et al. 2003) also assumes that two methods should have an equal variance to use that method.

The proposed model in this study is on the Bayesian inference, and it is also applicable to non-replicated measurements. The Bayesian inference allows offering

the more accurate, flexible, clear, and direct inference model using all available information (Bland and Altman 1999). Besides, Bayesian inference provides creativity in helpful directions because of flexibility. Also, it is simple in defining and interpreting (Boscardin and Gelman 1994; Hanneman 2018). More information can be incorporated in an inference when considering the necessity of specifying prior distributions. The information we gathered from the past we used as the prior information. Moreover, the Bayesian inference can be used with high accuracy even for small sample sizes. Another noteworthy reason nowadays is that Bayesian inference can be considered the breakthrough in computational algorithms and computing speed when performing very complex and realistic calculations. When dealing with the missing data, Bayesian methods allow for great flexibility (Kass and Wasserman 1996; Lin 1989). Based on the above reasons, Bayesian inference is a more suitable approach for method comparison studies dealing with heteroscedastic measurements.

The proposed model is the pioneer model using Bayesian inference to deal with the heteroscedastic method comparison data. The model is set up using the Bayesian theorem, and it calculates the posterior using likelihood, priors, and conditioning on observed data are rigid (Gardner and Altman 1986). Rather than removing heteroscedasticity from measurements, we are dealing with it using Bayesian inference, as it is more suitable and easy to interpret the results. The Bayesian computations for heteroscedastic linear models were used for method comparison as the first experience to deal with the heteroscedasticity measurements (Lim 2000; Lin et al. 2002). Also, we used a simulation study to examine and evaluate the performance of the proposed model. Finally, a practical illustration was used to verify the proposed model in method comparison data.

This article is organized as follows. Section 2 presents the proposed Heteroscedastic Bayesian Regression model for method comparison data. Section 3 explains the procedure to evaluate the agreement in method comparison data under the proposed heteroscedastic model, and section 4 explains the simulation study results. Section 5 shows an application of the proposed model using the gold particles

dataset. In Section 6, the conclusion describes the main findings and further discusses the limitations of the study.

2. BAYESIAN HETEROSCEDASTIC MODEL FOR METHOD COMPARISON

Let Y_i ; $i = 1, 2, \dots, n$ be the observed i^{th} measurement of the new method and X_i ; $i = 1, 2, \dots, n$ be the observed i^{th} measurement of the reference method. Y_i 's are assumed to be normally distributed with mean μ_i and variance σ_i^2 where the variance is not constant because it depends on the magnitude of measurements. The extended version of simple linear regression is used to deal with heteroscedastic measurements. The proposed heteroscedastic measurement method for method comparison is defined as below.

First, consider the homoscedastic model,

$$Y_i = \alpha + \beta * X_i + \varepsilon_i; i = 1, 2, \dots, n, \quad (1)$$

where $\mu_i = \alpha + \beta * X_i$ and $\varepsilon_i \sim N(0, \sigma^2)$.

$$Y_i \setminus \alpha, \beta, \sigma^2 \sim N(\alpha + \beta * X_i, \sigma^2)$$

where α and β are constants. The precision of this model can be expressed as, $\tau = 1/\sigma^2$.

The extended version of the above model for heteroscedastic model,

$$Y_i = \alpha + \beta * X_i + \varepsilon_i \text{ where } i = 1, 2, \dots, n \quad (2)$$

$$\mu_i = \alpha + \beta * X_i \text{ and } \varepsilon_i \sim N(0, \sigma_i^2)$$

$$Y_i \setminus \alpha, \beta, \sigma_i^2 \sim N(\alpha + \beta * X_i, \sigma_i^2)$$

where α and β , are constants. The precision of this model can be expressed as, $\tau = 1/(\sigma_i^2)$. Here X_i and Y_i are normally distributed. Rather than normality assumption, there are no other special assumptions because we used Bayesian inference for the proposed model. Also, we identify the function of X_i that can deal with the variance σ_i^2 and it depends on the dataset. Hence we should select the function of X_i according to the dataset.

In Bayesian implementation, it is crucial to identify the specific prior distribution, which leads to estimating the accurate posterior distribution. Let the prior distribution

be $p(\beta, \eta)$ for simplicity (Montenij et al. 2016). Then, the prior distribution can be written as below:

$$\begin{pmatrix} \beta \\ \eta \end{pmatrix} \sim N \left[\begin{pmatrix} b \\ g \end{pmatrix}, \begin{pmatrix} b & C \\ C' & G \end{pmatrix} \right],$$

Using Bayes theorem, we can obtain the posterior distribution as $\pi(\beta, \eta | \text{data}) \propto L(\beta, \eta)p(\beta, \eta)$ where $L(\beta, \eta)$ is the likelihood function under the normal distribution. According to the past literature, we have to use non-informative priors (Lin 2000). Hence, we must set up a different parameter setting and select the more appropriate priors related to our proposed model to give more accurate results.

3. EVALUATION OF THE AGREEMENT UNDER THE HETEROSCEDASTIC MODEL

Agreement evaluation is the main aim of the method comparison. Method comparison helps in decision making to select a more appropriate method. Concordance Correlation Coefficient (CCC) (Lin 1989; Lawrence et al 2007), Total Deviation Index (TDI), Intra-class Correlation Coefficient (ICC) are popular agreement measures used in method comparison (Chinchilli et al. 1996; Choudhary and Nagaraja 2007; Choudhary 2009; Choudhary and Yin 2010).

CCC is one of the agreement measures that takes values between -1 and +1. Higher values indicate a better agreement. The CCC measure presented by (Nawarathna and Choudhary 2015), defined as the,

$$CCC = \frac{2\rho\sigma_{1i}\sigma_{2i}}{(\mu_1 - \mu_2)^2 + \sigma_{1i}^2 + \sigma_{2i}^2} \quad (3)$$

According to equation (3), μ_1, μ_2 are the means of two measurements and σ_1, σ_2 are the respective variance's while ρ is the correlation coefficient of two measurements. The Heteroscedastic version of CCC is equation (4).

$$CCC = \frac{2 \text{cov}(Y_i, X_i) \sigma_{1i} \sigma_{2i}}{(\mu_1 - \mu_2)^2 + \sigma_{1i}^2 + \sigma_{2i}^2} \quad (4)$$

TDI measure indicates good agreement by small values. It is a non-negative value, and TDI is introduced by (Montenij et al. 2016). The general equation of TDI is equation (5).

$$TDI(\pi_0) = \tau \left\{ X_1^2 \left(\pi_0, \frac{\mu^2}{\tau^2} \right) \right\}^{1/2} \tag{5}$$

In the above equation, $X_1^2(\pi_0, \Delta)$ denotes the π_0^{th} percentile of a non-centralized Chi-Squared distribution with one degree of freedom. Generally $0.80 \leq \pi_0 \leq 0.9$, and for this calculation, π_0 taken as 0.95. Its heteroscedastic version with the proposed model can be obtained as,

$$TDI(\pi_0) = \tau \left\{ X_1^2 \left(\pi_0, \frac{(\alpha + \beta X_i)^2}{(\eta + Y\sqrt{X_i})^2} \right) \right\}^{1/2} \tag{6}$$

ICC is another agreement measure that takes values between 0 and +1. Moreover, higher values indicate good agreement. ICC value can be expressed as,

$$ICC = \frac{\sigma_{12}}{(\bar{Y} - \bar{X})^2 + \sigma_{1i}^2 + \sigma_{2i}^2 - \frac{\sigma_{Di}^2}{n}} \tag{7}$$

σ_{12} is the covariance of the two measurements and \bar{Y}, \bar{X} indicate the mean of two measurements. Also $\sigma_{1i}^2, \sigma_{2i}^2$ and σ_{Di}^2 are the variance of two methods and the difference between two methods, respectively. This measure also can be presented in the heteroscedastic version in equation (8).

$$ICC = \frac{cov(Y_i, X_i)}{(\bar{Y} - \bar{X})^2 + \sigma_{1i}^2 + \sigma_{2i}^2 - \frac{(\eta + Y\sqrt{X_i})^2}{n}} \tag{8}$$

Credible Interval (CI) (Peter 2001) is a measure that gives the interval, which lies the value of a parameter. When estimating the unknown parameters, CI describes and summarizes the uncertainty of the unknown parameter. CI is just the range or interval containing a particular percentage of probable values in the domain of Posterior distribution. Generally $CI = \text{sample statistics} \pm \text{margin of error}$. The product of the critical value (z) and standard error of the point estimator gives the margin of error using the standard normal table. The critical value can be calculated. Standard error =

$\frac{\sigma_i}{\sqrt{n}}$ where σ_i is the standard deviation of the point estimate and n is the sample size.

The following equations illustrate the CI for all parameters in the proposed model.

$$CI = \hat{\alpha} \pm Z_{\frac{\alpha}{2}} \left(\frac{(\eta_0 + \eta_1 \sqrt{X_i})}{\sqrt{n}} \right) \quad (9)$$

$$CI = \hat{\beta} \pm Z_{\frac{\alpha}{2}} \left(\frac{(\eta_0 + \eta_1 \sqrt{X_i})}{\sqrt{n}} \right) \quad (10)$$

$$CI = \hat{\eta}_0 \pm Z_{\frac{\alpha}{2}} \left(\frac{(\eta_0 + \eta_1 \sqrt{X_i})}{\sqrt{n}} \right) \quad (11)$$

$$CI = \hat{\eta}_1 \pm Z_{\frac{\alpha}{2}} \left(\frac{(\eta_0 + \eta_1 \sqrt{X_i})}{\sqrt{n}} \right) \quad (12)$$

Here $Z_{\frac{\alpha}{2}}$ is taken from the normal distribution while α is the confidence coefficient.

4. SIMULATION STUDY

This section is used to evaluate and compare the accuracy of the proposed model under different settings and different sample sizes. The main focus of this section is to assess the characteristics of the proposed model using the Monte Carlo simulation (Linnet 1998; Montenij et al. 2016). Results of the model under different conditions are foreseen using simulation and representing the model behaviour more accurately. Just another Gibbs sampler (jags) programme was used for simulation because jags provides a helpful Bayesian modelling platform (Nawarathna and Choudhary 2013). The model fitting and analysis was performed by R 3.6.1 statistical software.

This study selects six different settings to evaluate the model behaviour. Simulation settings are set according to the strength of the variance and agreement. Besides, each of these settings also run under different sample sizes 1000 times. This simulation process can be divided into two steps. Under the first step, we calculate the point estimator for each parameter and credible intervals, coverage probabilities for each trial to evaluate the model's accuracy. In the second step, six different settings and sample sizes are used to assess the proposed model's performance and behaviour by calculating credible interval and coverage probabilities (cp) with error values. Initial values of the parameter values are given in table I.

Table I: Initial values for all parameters.

Parameter	α	β	η_0	η_1
Initial value	0.13	0.576	-1.551	-0.523

As the first part of the simulation, average values, coverage probabilities, credible intervals, and error values are calculated for each run for all the point estimates to measure the accuracy of the proposed model. The simulation results imply that all the data points perfectly match the proposed model with acceptable coverage probabilities with minimum error values.

The coverage probabilities of α , β , η_0 and η_1 are 0.975, 0.95, 0.96, and 0.965, respectively. These coverage probabilities imply that each estimator's probability lies between appropriate credible intervals is higher than 0.95, and all probabilities are in an acceptable range. In this simulation, we calculated the four error values to get an idea about the accuracy. Four error measurements are calculated, namely Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE), and those values are 0.034, 0.1829, 0.4316 and 0.4811 respectively. All the error values are relatively low, implying that the proposed model has high accuracy. These values also indicate the good significance of the proposed model.

It is necessary to identify the characteristics of the model with the different settings under the different sample sizes. We have selected six different settings for the second part of the simulation. When selecting these different settings, variance changes and the strength of the agreement are considered. We want to identify the model behavior for selecting these different settings with the changes of heteroscedastic parameters and changes of the agreement. Moreover, we verify the model behavior with different sample sizes. Our second simulation step was carried out to verify the characteristics under sample sizes of 20, 50, 70, 100, 500 and 1000. Values of the samples are estimated by the proposed model using selected priors.

The simulation was carried out for all samples sizes under each setting. Simulation results for different settings under different sample sizes are summarized

in Table II with respective error values. Moreover, Figures 1 & 2 represent the graphical representations of simulation results of error value and coverage probability changes according to the sample sizes. For each setting, simulation is run 1000 times to ensure the accuracy of the results.

In all the settings, sample size 500 shows the good coverage probabilities for all parameters. Further, coverage probabilities have higher values than others in all six settings when the agreement is high. Coverage probabilities of the low variance related settings have reasonable values than other settings. The error values for all sample sizes are shallow, and when variance is high, error values are lower in many settings. According to the results, the best sample size for this proposed model is 500. The model plays well when the measurements are in high agreement. In this way, simulation verifies the characteristics of the proposed model more clearly and accurately.

Table II: Summary simulation results for different settings and sample sizes.

Sample size	Settings	α	CP of α	β	CP of β	η_0	CP of η_0	η_1	CP of η_1	MSE	RMSE	SSE	MAPE	SMAPE
20	1	0.038	0.97	0.015	0.12	-0.072	0.11	-0.032	0.15	0.078	0.274	0.312	4.866	1.196
	2	0.038	0.95	0.015	0.12	-0.072	0.13	-0.032	0.12	0.076	0.272	0.304	4.821	1.166
	3	0.038	0.96	0.015	0.14	-0.072	0.11	-0.032	0.13	0.089	0.28	0.355	5.352	1.239
	4	0.039	0.97	0.015	0.13	-0.072	0.12	-0.031	0.11	0.066	0.254	0.264	4.575	1.186
	5	0.04	0.95	0.015	0.11	-0.072	0.13	-0.032	0.13	0.103	0.245	0.24	4.022	1.141
	6	0.113	0.96	0.632	0.12	-2.338	0.14	0.14	-0.305	0.12	0.071	0.26	0.282	4.796
50	1	0.109	0.975	0.069	0.21	-0.17	0.2	-0.065	0.19	0.205	0.399	1.638	1.813	0.709
	2	0.113	0.98	0.078	0.16	-0.165	0.17	-0.060	0.18	0.4	0.245	0.555	1.204	0.641
	3	0.111	0.95	0.077	0.15	-0.166	0.16	-0.06	0.17	0.088	8.231	0.705	1.45	0.695
	4	0.112	0.96	0.075	0.13	-0.168	0.13	-0.063	0.14	0.138	0.341	1.102	1.384	0.632
	5	0.114	0.97	0.077	0.15	-0.165	0.14	-0.06	0.15	0.067	0.243	0.539	1.169	0.62
	6	0.108	0.96	0.069	0.12	-0.17	0.13	-0.064	0.13	0.201	0.428	1.605	1.885	0.78
70	1	0.124	0.98	0.099	0.12	-0.239	0.14	-0.092	0.4	0.156	0.382	1.873	1.124	0.613
	2	0.125	0.975	0.103	0.11	-0.239	0.12	-0.17	0.11	0.124	0.320	1.489	0.972	0.542
	3	0.123	0.97	0.103	0.975	-0.239	0.98	-0.092	0.97	0.131	0.349	1.575	1.105	0.580
	4	0.125	0.98	0.185	0.12	-0.238	0.12	-0.092	0.12	0.119	0.315	1.426	1.010	0.570
	5	0.126	0.95	0.104	0.11	-0.237	0.11	-0.09	0.14	0.099	0.290	1.336	0.943	0.570
	6	0.125	0.96	0.106	0.13	-0.236	0.12	-0.089	0.13	0.082	0.267	0.979	0.880	0.571

Sample size	Settings	α	CP of α	β	CP of β	η_0	CP of η_0	η_1	CP of η_1	MSE	RMSE	SSE	MAPE	SMAPE
100	1	0.146	0.98	0.115	0.12	-0.347	0.5	-0.141	0.11	0.104	0.31	1.974	0.858	0.542
	2	0.148	0.96	0.116	0.11	-0.347	0.12	-0.142	0.12	0.102	0.291	1.939	0.826	0.532
	3	0.146	0.95	0.116	0.12	-0.348	0.12	-0.142	0.13	0.114	0.334	2.174	0.86	0.539
	4	0.147	0.97	0.115	0.13	-0.346	0.11	-0.139	0.12	0.091	0.289	1.742	0.869	0.575
	5	0.147	0.95	0.115	0.11	-0.347	0.13	-0.142	0.11	0.105	0.311	1.996	0.849	0.556
	6	0.148	0.98	0.12	0.12	-0.345	0.12	-0.14	0.13	0.085	0.282	1.607	0.8	0.523
500	1	0.136	0.965	0.559	0.9	-1.534	0.97	-0.494	0.95	0.030	0.168	2.995	0.409	0.482
	2	0.134	0.97	0.559	0.96	-0.541	0.98	-0.494	0.975	0.032	0.177	3.159	0.426	0.497
	3	0.134	0.975	0.567	0.95	-0.231	0.96	-0.489	0.98	0.03	0.171	2.958	0.417	0.489
	4	0.134	0.96	0.56	0.97	-1.534	0.95	-0.483	0.96	0.03	0.171	0.95	0.421	0.484
	5	0.132	0.97	0.564	0.98	-1.536	0.97	-0.495	0.97	0.031	0.174	3.080	1.314	0.469
	6	0.134	0.98	0.569	0.95	-0.541	0.96	-0.494	0.95	0.031	0.176	3.129	0.41	0.473
1000	1	0.126	0.96	0.534	0.9	-2.024	0.5	-0.386	0.8	0.078	0.253	1.913	0.701	0.52
	2	0.116	0.975	0.621	0.7	-2.321	0.13	-0.3	0.12	0.028	0.167	5.556	0.389	0.443
	3	0.115	0.8	0.625	0.5	-0.351	0.12	-0.311	0.11	0.031	0.177	6.243	0.389	0.438
	4	0.115	0.9	0.626	0.5	-2.327	0.12	-0.281	0.3	0.036	0.189	7.191	0.45	0.445
	5	0.039	0.9	0.015	0.6	-0.072	0.13	-0.032	0.11	0.028	0.164	5.398	0.39	0.439
	6	0.117	0.975	0.621	0.6	-2.330	0.12	-0.312	0.13	0.031	0.176	6.217	0.395	0.452

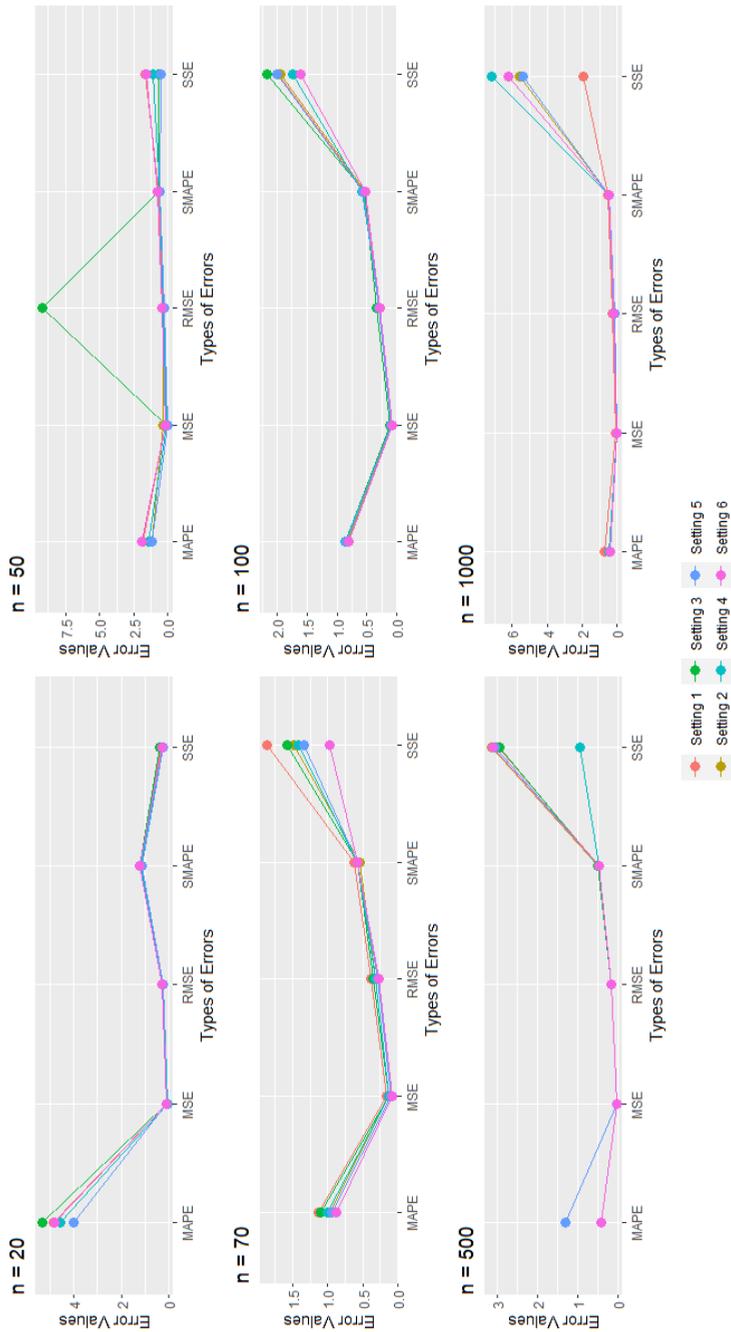


Fig 1: Error values of each setting according to the sample sizes; n = 20, 50, 70, 100, 500 and 1000.

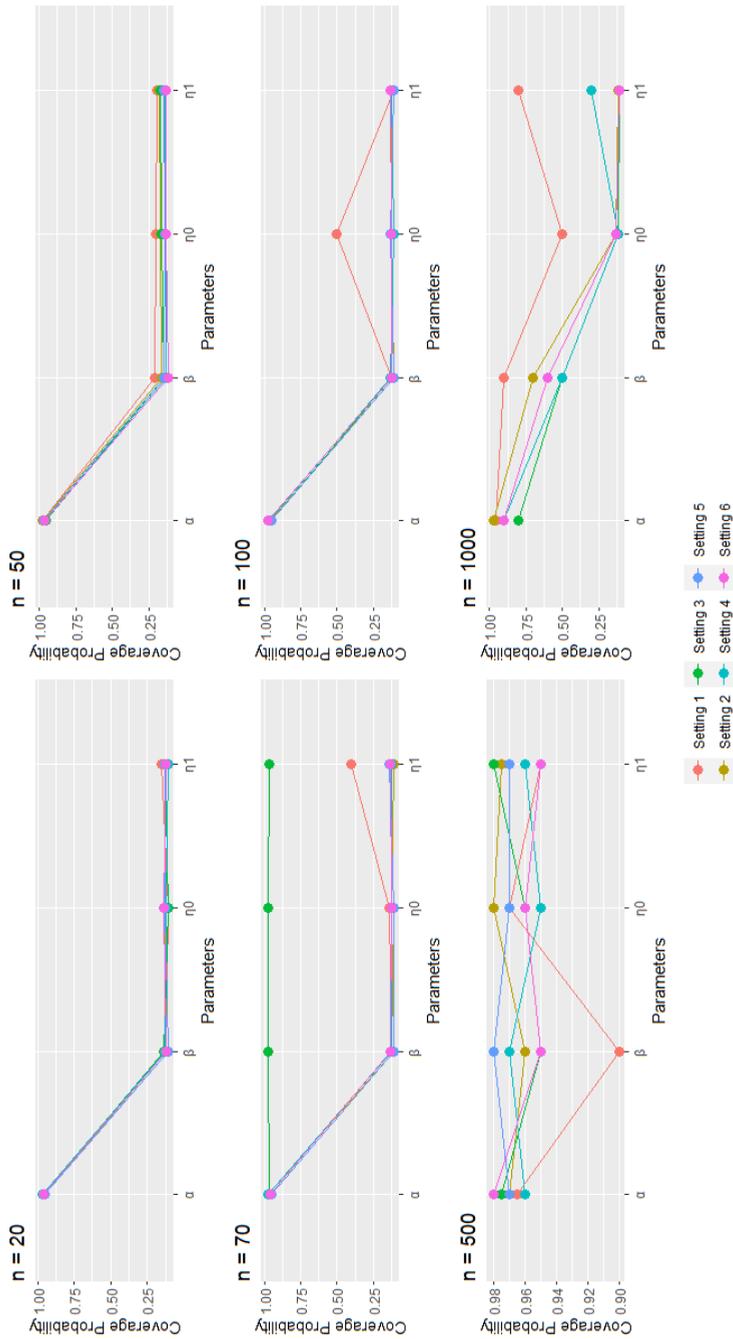


Fig 2: Coverage Probabilities of Parameters according to sample sizes; $n = 20, 50, 70, 100, 500$ and 1000 .

5. APPLICATION

This section explains the practical viewpoint of the proposed model. For this purpose, the gold particles dataset was selected from (Tim et al. 2019). The dataset consists of gold particles in copper deposits with 501 subjects. Also, it involves two methods called Classical Fire Assay (CFA) and Screen Fire Assay (SFA), which are used to check the content of gold particles. The existing classical method serves as a reference method, and the screen fire assay method serves as a test method for this analysis. The method comparison aims to check that the measurements of the two methods agree well and those two methods can be comparable. A small number of gold particles are included in the copper deposits deposit, and the chemical laboratory collects those samples. Hence, a more accurate method is essential. We evaluate the agreement using the proposed model and check whether the two methods have a satisfactory agreement to interchange the two methods. The trellis plot of the gold particles dataset displayed in Figure 3 shows how data appears in two methods by two colours.

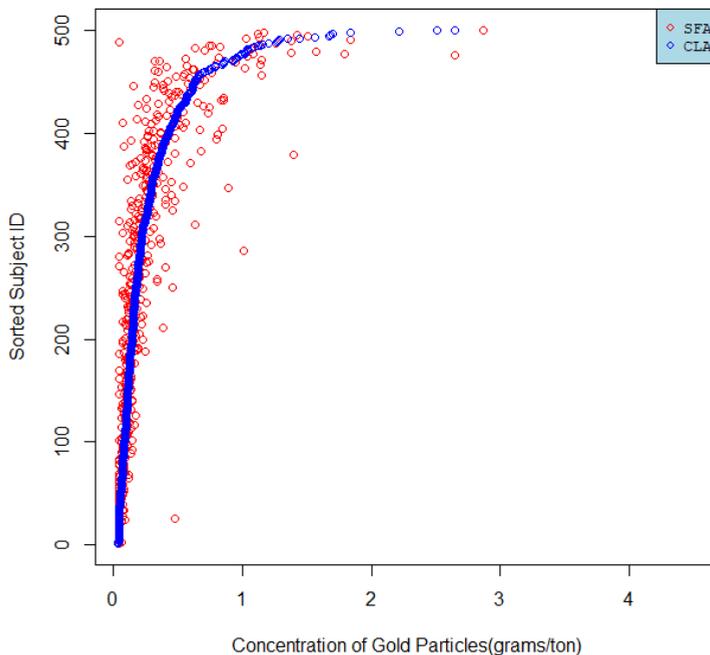


Fig 3: Trellis plot of gold particles dataset.

The above plot represents the measurements of two methods separately by colour, and CFA shows a higher within-subject variation than the SFA measurements. Moreover, the plot indicates that the variation of both methods increases with the magnitude of the measurements.

Then, we select the suitable version of the heteroscedastic parameter for this dataset. We used several linear and non-linear equations and calculated the error values to measure the accuracy. Finally, the following non-linear equation was established for gold particles data because it gives the minimum error values with higher accuracy. Hence we selected this equation to deal with the heteroscedasticity of this proposed model.

$$\sigma_i^2 = \exp(\eta_0 + \eta_1 \sqrt{X_i})$$

Further, priors for the proposed model can be mentioned as follows.

$$\alpha \sim N(0, 100)$$

$$\beta \sim N(0, 100)$$

$$\eta_0 \sim N(0, 100)$$

$$\eta_1 \sim N(0, 100)$$

To check the accuracy of the proposed model, the above priors are used. Summary results of model fitting for gold particles data are summarized in Table III. MSE, RMSE, MAPE, and SMAPE values are calculated to examine the accuracy of the fitted proposed model. Also, CIs are calculated for each parameter, implying the range of the values for parameters and, according to table values, estimated α near to zero and estimated β value near to one.

	α	β	η_0	η_1
Estimated value	0.13	0.576	-1.551	-0.523
Lower Limit	0.084	0.502	-1.723	-0.691
Upper Limit	0.175	0.648	-1.378	-0.354

Table III: summary results of model fitting for gold particles data.

According to Table III, a better agreement of two methods is implied. To examine the accuracy of the fitted proposed model, MSE, RMSE, MAPE, and SMAPE values are calculated, and the estimated values are 0.0444, 0.2107, 0.4203, and 0.4716, respectively. Hence, all error values are relatively low, and the fitted model has higher accuracy.

Moreover, agreement measures are calculated to check whether the two methods can be interchanged. The estimated CCC, TDI, and ICC values are 0.8387, 0.055 and 0.4194, respectively. All the values are verified that the two methods have a good agreement to use interchangeably.

6. CONCLUSION

This paper proposes the Bayesian model to evaluate the agreement between two methods when the variability of measurements depends on the magnitude. The proposed model fits well with low error values and high coverage probabilities. Simulation results show that all parameters have 0.95 or more than coverage probability. Those results imply the best fit of the proposed model. Moreover, simulation results indicate that this model is more suitable for sample size 500 and not performing well with sample sizes less than 20. Moreover, the proposed model works well when high heteroscedasticity is present with high accuracy. One limitation of the proposed model is that it does not fit well with small sample sizes.

ACKNOWLEDGMENTS

Many thanks to my supervisors, Postgraduate Institute of Science, University of Peradeniya and my family.

REFERENCES

- [1] ALTMAN, D.G., AND BLAND, J.M. 1983. Measurement in medicine: the analysis of method comparison studies. *The Statistician*. 32, 307-317.
- [2] ANDREW, G., CARLIN, J.B., STERN, H.S., et al. 2004. Bayesian Data Analysis, 2nd edition. Chapman & Hall/CRC.
- [3] ARAVIND, K.R., AND NAWARATHNA, L.S. 2017. A statistical method for assessing agreement between two methods of heteroscedastic clinical measurements using the copula method. *J Med Stat Inform*. 5:3. 10.7243/2053-7662-5-3
- [4] BILIĆ-ZULLE, L. 2011. Comparison of methods: Passing and Bablok regression. *Biochemia Medica*. 21(1):49-52. 10.11613/BM.2011.010.
- [5] BLAND, J.M., AND ALTMAN, D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 307—310.
- [6] BLAND, J.M., AND ALTMAN, D.G. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*. 8(2): 135–160. 10.1177/096228029900800204.
- [7] BOSCARDIN, W.J., AND GELMAN, A. 1994. Bayesian Computation for Parametric Models of Heteroscedasticity in the Linear Model. *In Advances in Econometrics*. 11, A87- 10.
- [8] BOSCARDIN, W.J., GELMAN, A. 1996. Bayesian Computation for parametric Models of Heteroscedasticity in the Linear Model. *In Advances in Econometrics*, Volume 11 (Part A), edited by Hill, R. C. Connecticut: JAI Press Inc.
- [9] CARSTENSEN, B. 2010. Comparing methods of measurement: Extending the LoA by regression. *Statistics in Medicine*. 29: 401-410. doi:10.1002/sim.3769
- [10] CHINCHILLI, V.M., MARTEL, J.K., KUMANYIKA, S., AND LLOYD, T. 1996. A weighted concordance correlation coefficient for repeated measurement designs. *Biometrics*. 52:341–353.
- [11] CHOUDHARY, P.K., AND NAGARAJA, H.N. 2007. Tests for assessment of agreement using probability criteria. *Journal of Statistical Planning and Inference*.137:279–290.
- [12] CHOUDHARY, P.K. 2009. Interrater agreement. *In Methods and Applications of Statistics in the Life and Health Sciences*, BALAKRISHNAN N et. al. (ed.). John Wiley: New York. 461–480.
- [13] CHOUDHARY, P.K., AND YIN, K. 2010. Bayesian and Frequentist Methodologies for Analyzing Method Comparison Studies With Multiple Methods. *Statistics in Biopharmaceutical Research*. 2:122-132. 10.1198/sbr.2010.08096.
- [14] DE WAAL, E.E., WAPPLER, F., AND BUHRE, W.F. 2009. Cardiac output monitoring. *Curr Opin Anaesthesiol*. 22: 71–7.
- [15] EDILBERTO, C., AND JORGE, A. 2009. Regression Models with Heteroscedasticity using Bayesian Approach. *Revista Colombiana de Estadística*.32
- [16] GALEA-ROJAS, M., DE CASTILHO, M.V., AND BOLFARINE, H., et al. 2003. Detection of analytical bias. *Analyst*. 128: 1073–1081.
- [17] GARDNER, M.J., AND ALTMAN, D.G. 1986. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*. 292:746–50.
- [18] HANNEMAN, S.K. 2018. Design, analysis, and interpretation of method-comparison studies. *AACN advanced critical care*. 19(2): 223–234. 10.1097/01.AACN.0000318125.41512.a3.
- [19] KASS, R.E., AND WASSERMAN, L. 1996. The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association*. 91:435, 1343-1370. 10.1080/01621459.1996.10477003
- [20] LIN, L., HEDAYAT, A.S., WU, W. 2007. A Unified Approach for Assessing Agreement for Continuous and Categorical Data. *Journal of Biopharmaceutical Statistics*, 17:4, 629-652, DOI: 10.1080/10543400701376498
- [21] LIN, L. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45:255–268. Corrections: 2000,56, 324–325.
- [22] LIN, L. 2000. Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Statistics in Medicine*. 19:255–270.

-
- [23] LIN, L.I., HEDAYAT, A.S., SINHA, B., AND YANG, M. 2002. Statistical methods in assessing agreement: models, issues, and tools. *Journal of the American Statistical Association*.97:257–270.
- [24] LINNET, K. 1998. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clinical Chemistry*. 44(5):1024-1031.
- [25] MONTENIJ, L.J., BUHRE, W.F., JANSEN, J.R. et al. 2016. methodology of method comparison studies evaluating the validity of cardiac output monitors: a stepwise approach and checklist. *British Journal of Anaesthesia*. 116 (6): 750–8.
- [26] NAWARATHNA, L.S., CHOUDHARY, P.K. 2013. Measuring Agreement in Method Comparison Studies with Heteroscedastic Measurements. *Statistics in Medicine*. 32(29): 5156-5171. <https://doi.org/10.1002/sim.5955>
- [27] NAWARATHNA, L.S., CHOUDHARY, P. 2015. A heteroscedastic measurement error model for method comparison data with replicate measurements. *Statistics in Medicine*. 34(7), 1242-1258.
- [28] O'HAGAN, ANTHONY, FORSTER, et al 2004. *Reading in Bayesian Inference*. Volume 2b of Kendall's Advanced Theory of Statistics, second edition. Arnold, London.
- [29] PETER, B. 2001. A Brief Introduction to Monte Carlo Simulation. *Clinical pharmacokinetics*.40: 15-22. 10.2165/00003088-200140010-00002.
- [30] PEYTON, P.J., AND CHONG, S.W. 2010. Minimally invasive measurement of cardiac output during surgery and critical care: a meta-analysis of accuracy and precision. *Anesthesiology*. 113:1220–35.
- [31] STUDENMUND, A.H. 1991. *Reading using Econometrics: A practical guide*.2nd Ed. ISBN 0-673-52125-7.
- [32] TIM, P.M., IAN, R.W., AND MICHAEL, J.C. 2019. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 38:2074–2102. doi.10.1002/sim.8086

S. M. M. Lakmali,
Postgraduate Institute of Science,
University of Peradeniya, Peradeniya, Sri Lanka.

Lakshika S. Nawarathna,
Department of Statistics and Computer Science,
Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.
Email: lakshikas@pdn.ac.lk

P. Wijekoon,
Department of Statistics and Computer Science,
University of Peradeniya, Peradeniya, Sri Lanka.

A new generalized transmuted distribution

S. A. WANI AND S. A. DAR

Abstract

We introduced Transmuted another Two-Parameter Sujatha Distribution by using Quadratic Rank Transmutation Map technique. Various necessary statistical properties of Transmuted another Two-Parameter Sujatha Distribution are obtained. The reliability measures of proposed model are also derived and model parameters are estimated by using maximum likelihood estimation method. The significance of transmuted parameter has been tested by using likelihood ratio statistic. Finally, an application to real data sets is presented to examine the significance of newly introduced model by computing Kolmogorov statistic, p-value, AIC, BIC, AICC, HQIC.

General Terms: Method Comparison, Simulation, Bayesian Inference.

Mathematics Subject Classification: 60E05, 62F10

Keywords: Quadratic Rank Transmutation Map technique, Simulation study, Transmuted parameter, Statistical properties and Maximum likelihood estimation.

1. INTRODUCTION

Data analysts and researchers fit appropriate model to various real life data sets depending on the nature of the real life data among other things. A model with large number of parameters brings more flexibility and covers more variation from the data. There are many techniques for obtaining generalized probability models. One of the techniques for adding extra parameter to the existing models is Quadratic Rank Transmutation Map technique (QRTM). Data analysts make use of Quadratic Rank Transmutation Map technique for analyzing very complex data. Shaw and Buckley (2007) introduced Quadratic Rank Transmutation Map technique for generalization of classical probability models [1]. Aryal and Tsokos (2009) worked on transmuted generalized extreme value distribution and studied its applications and properties [2]. Hassan, Wani & Para (2018) introduced three parameter Quasi Lindley distribution by using weighting technique and studied various properties of that model [3]. Hassan, Wani, Shafi and Sheikh (2020) introduced Lindley-Quasi Xgamma Distribution (LQXD) and studied its applications along with properties [4] Merovci (2013) developed transmuted Rayleigh distribution and studied its

10.2478/jamsi-2022-0013

©Sameer Ahmad Wani, Showkat Ahmad Dar

This is an open access article licensed under the Creative Commons

Attribution Licence (<http://creativecommons.org/licenses/by/4.0>).



necessary properties and applications [5]. Para and Jan (2018) formulated transmuted inverse log-logistic distribution and obtained its various characteristic properties [6]. Haq (2016) studied transmuted exponentiated inverse Rayleigh distribution and obtained its various properties [7]. Hassan, Wani, Shafi (2020) introduced Poisson Pranav distribution and obtained its various mathematical properties along with obtaining applications of the proposed model [8].

Here we have incorporated an extra parameter known as transmuted parameter to Another Two-Parameter Sujatha distribution [9].

A continuous random variable X is said to follow another Two-Parameter Sujatha Distribution (ATPSD) if its probability density function is of the form

$$g_b(x) = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \quad x > 0, \theta > 0, \alpha \geq 0 \quad (1.1)$$

And is denoted by $X \sim \text{ATPSD}(\alpha, \theta)$

The cumulative distribution function of Another Two-Parameter Sujatha Distribution is given by

$$G_b(x) = 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \quad x > 0, \theta > 0, \alpha \geq 0 \quad (1.2)$$

By using Quadratic Rank Transmutation Map technique. The c.d.f of transmuted model $F_\tau(x)$ is of the below form

$$F_\tau(x) = (1 + \lambda)G_b(x) - \lambda[G_b(x)]^2, \quad -1 \leq \lambda \leq 1,$$

Which on differentiation yields the p.d.f $f_\tau(x)$ of transmuted model as

$$f_\tau(x) = g_b(x)[1 + \lambda - 2\lambda G_b(x)]$$

Where $g_b(x)$ and $G_b(x)$ are p.d.f and c.d.f of base model respectively.

2. TRANSMUTED ANOTHER TWO-PARAMETER SUJATHA DISTRIBUTION (TATPSD)

A non-negative random variable X is said to have Transmuted another Two-Parameter Sujatha Distribution (TATPSD) if its cumulative distribution function $F_{\tau}(x)$ is obtained as

$$F_{\tau}(x, \alpha, \theta, \lambda) = (1 + \lambda)G_b(x) - \lambda[G_b(x)]^2, -1 \leq \lambda \leq 1,$$

$$F_{\tau}(x, \alpha, \theta, \lambda) = G_b(x)\{1 + \lambda(1 - G_b(x))\} \tag{2.1}$$

Putting the value of $G_b(x)$ from equation (1.2) in equation (2.1) we get

$$F_{\tau}(x, \alpha, \theta, \lambda) = \left\{1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)}\right] e^{-\theta x}\right\} \left\{(1 + \lambda) - \lambda \left\{1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)}\right] e^{-\theta x}\right\}\right\} \tag{2.2}$$

The graphs of c.d.f of Transmuted Another Two-Parameter Sujatha Distribution are given below:

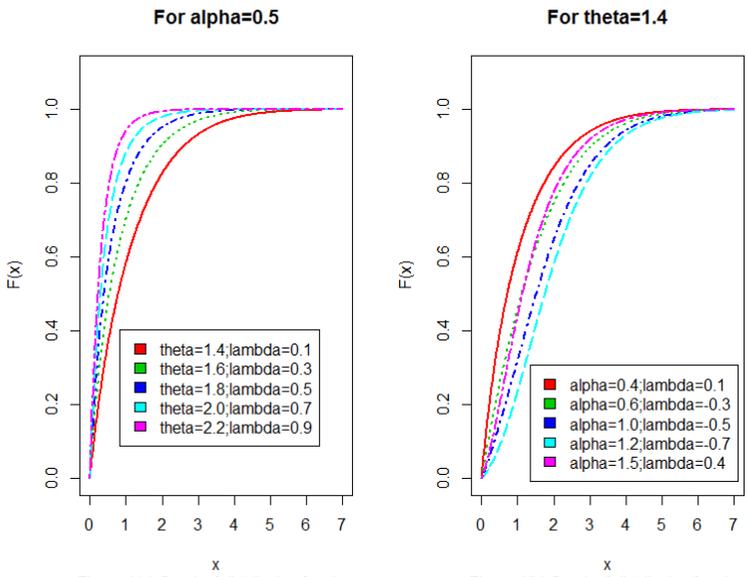


Figure 1(a) Graph of distribution function

Figure 1(b) Graph of distribution function

Where $G_b(x)$ is the c.d.f of base distribution (Another Two-Parameter Sujatha Distribution). For $\lambda = 0$ in (2.1) we get the base distribution.

As can be seen from Figures 1(a) and 1(b) c.d.f starts at zero and ends with one for different parameter values which shows that c.d.f is accurate.

The corresponding probability density function of Transmuted Another Two-Parameter Sujatha Distribution $f_{\tau}(x, \alpha, \theta, \lambda)$ is obtained by differentiating (2.1) as

$$f_{\tau}(x, \alpha, \theta, \lambda) = g_b(x)[1 + \lambda - 2\lambda G_b(x)] \quad (2.3)$$

Putting the value of $g_b(x)$ and $G_b(x)$ in equation (2.3)

$$f_{\tau}(x, \alpha, \theta, \lambda) = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \quad (2.4)$$

$$x > 0, \theta > 0 \text{ and } \alpha > 0, -1 \leq \lambda \leq 1$$

Where α, θ & λ are shape, scale and transmuted parameters respectively.

Which is the p.d.f of Transmuted another Two-Parameter Sujatha distribution.

Where $g_b(x)$ and $G_b(x)$ are p.d.f and c.d.f of base model respectively.

As can be seen from figure 2(a) and 2(b) the Transmuted another Two-Parameter Sujatha Distribution is positively skewed as is clear from graphs given below:

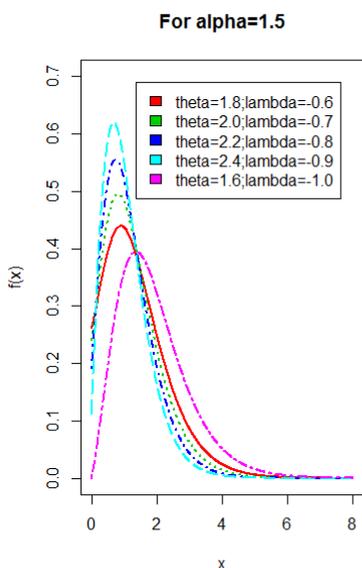


Figure 2(a) Graph of Probability Density Function

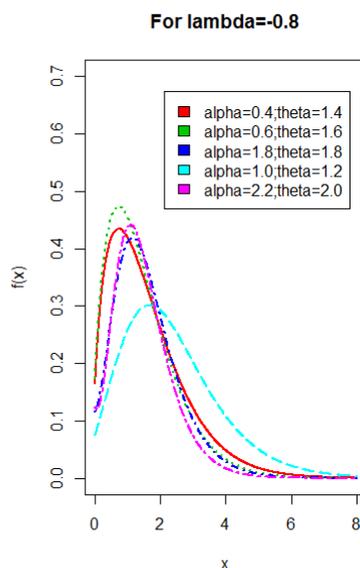


Figure 2(b) Graph of Probability Density Function

3. SPECIAL CASES

CASE 1: If we put $\lambda = 0$, then Transmuted another Two-Parameter Sujatha distribution (2.4) reduces to Another Two-Parameter Sujatha distribution with shape parameter α & scale parameter θ and having probability density function as:

$$f(x) = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \quad x > 0, \theta > 0, \alpha \geq 0$$

CASE 2: If we put $\alpha = 0, \lambda = 0$, then Transmuted another Two-Parameter Sujatha distribution (2.4) reduces to Exponential distribution having probability density function as:

$$f(x) = \theta e^{-\theta x} \quad x > 0, \theta > 0$$

4. RELIABILITY ANALYSIS

We explored survival function, hazard rate and reverse hazard rate of the proposed transmuted another Two-Parameter Sujatha distribution in this segment of paper.

4.1 Reliability function $R(x)$

The reliability function or survival function $R(x, \alpha, \theta, \lambda)$ is the measure of chance that a system survives beyond a specified time (t).

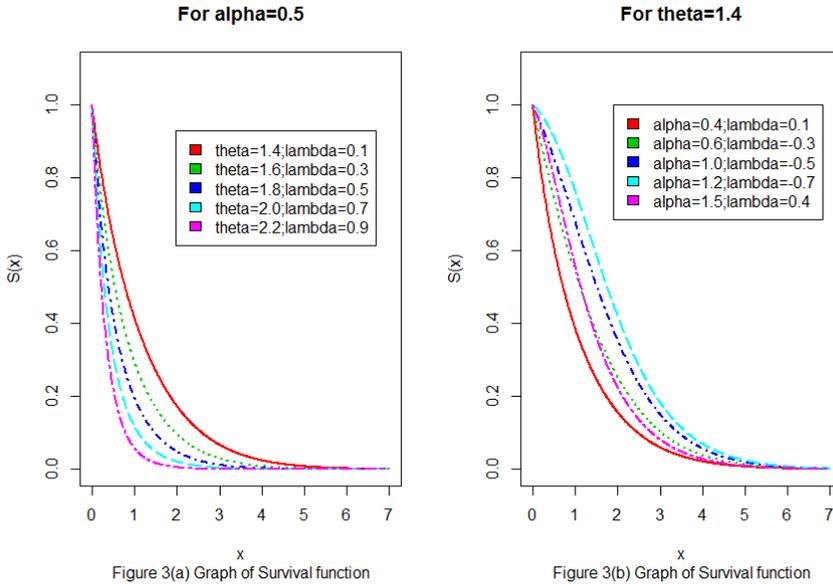
Mathematically

$$R(x, \alpha, \theta, \lambda) = P(X > t)$$

It can be obtained as complement of the cumulative distribution function of the model. The reliability function or the survival function of Transmuted another Two-Parameter Sujatha distribution is calculated as:

$$R(x, \alpha, \theta, \lambda) = 1 - \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\}$$

The graphs of reliability function are given below:



As can be seen from Figures 3(a) and 3(b) survival function starts at one and ends with zero for different parameter values which shows that survival function is accurate.

4.2 Hazard function

The hazard function of TATPSD is obtained as

$$\begin{aligned}
 H.R = h(x; \alpha, \theta, \lambda) &= \frac{f(x, \alpha, \theta, \lambda)}{R(x, \alpha, \theta, \lambda)} \\
 &= \frac{\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\}}{1 - \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left[1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\}}
 \end{aligned}$$

The graph of hazard rate of Transmuted another Two-Parameter Sujatha distribution is given below

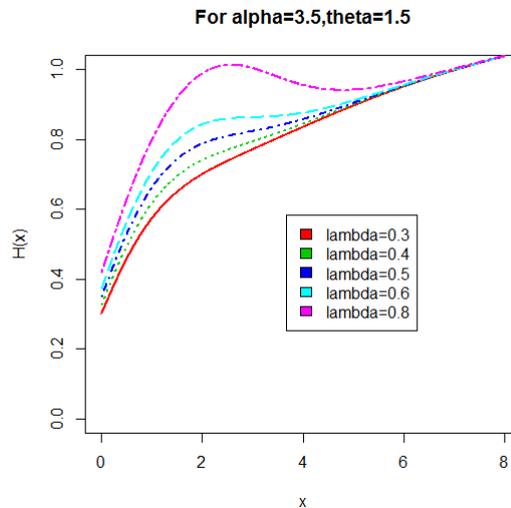


Figure 4 Graph of Hazard Function

The above graph of hazard rate of TATPSD reveals increasing, decreasing and constant hazard rate for different values of transmuted parameter. The shape of hazard rate graph is bathtub showing the flexibility of proposed model and greater applicability in real life.

4.3 Reverse hazard rate

The reverse hazard rate of the TATPSD is given as:

$$R.H.R = h(x, \alpha, \theta, \lambda) = \frac{f(x, \alpha, \theta, \lambda)}{F(x, \alpha, \theta, \lambda)}$$

$$= \frac{\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\}}{\left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left[1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\}}$$

5. STATISTICAL PROPERTIES

In this section we have obtained the different structural and statistical properties of the proposed Transmuted another Two-Parameter Sujatha model. These include moments, moment generating function and characteristic function

5.1 Moments

Suppose X is a random variable following Transmuted another Two-Parameter Sujatha distribution with parameters α, θ and λ . Then the r^{th} moment about origin for TATPSD is given by

$$\begin{aligned} \mu_r' &= E(X^r) = \int_0^\infty x^r f(x, \alpha, \theta, \lambda) dx && r=1, 2, 3 \dots n \\ &= \int_0^\infty x^r \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} dx \\ \mu_r' &= \left(\frac{r!}{\theta^r(\theta^2 + \alpha\theta + 2\alpha)} \right. \\ &\quad \left. \left\{ (1 - \lambda)(\theta^2 + (r + 1)\alpha\theta + \alpha(r + 1)(r + 2)) + \right. \right. \\ &\quad \left. \left. 2\lambda \left\{ \frac{1}{2^{r+3}} (4\theta^2 + 2\alpha\theta(r + 1) + \alpha(r + 1)(r + 2)) + \frac{1}{(\theta^2 + \alpha\theta + 2\alpha)} \right. \right. \right. \\ &\quad \left. \left. \left. \frac{\alpha(r+1)(r+2)}{2^{r+5}} (4\theta^2 + 2\alpha\theta(r + 3) + \alpha(r + 3)(r + 4)) + \frac{\alpha(\alpha\theta + 2)(r+1)}{2^{r+4}} \right\} \right\} \right) \end{aligned} \quad (5.1.1)$$

5.2 Moment generating function and characteristic function of TATPSD

We will derive moment generating function and characteristic function of TATPSD in this section of paper.

THEOREM 5.2.1: If X has the TATPSD(θ, α, λ), then the moment generating function $M_X(t)$ and characteristic generating function $\varphi_X(t)$ are

$$M_X(t) = \left(\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left\{ (1 - \lambda) \left(\frac{1}{\theta - t} + \frac{\alpha}{(\theta - t)^2} + \frac{2\alpha}{(\theta - t)^3} \right) + 2\lambda \left\{ \left(\frac{1}{(2\theta - t)} + \frac{\alpha}{(2\theta - t)^2} + \frac{2\alpha}{(2\theta - t)^3} \right) + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} \right\} \left\{ \theta \left(\frac{2}{(2\theta - t)^3} + \frac{6\alpha}{(2\theta - t)^4} + \frac{24\alpha}{(2\theta - t)^5} \right) + (\alpha\theta + 2) \left(\frac{1}{(2\theta - t)^2} + \frac{2\alpha}{(2\theta - t)^3} + \frac{6\alpha}{(2\theta - t)^4} \right) \right\} \right) \right)$$

and

$$\varphi_X(t) = \left(\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left\{ (1-\lambda) \left(\frac{1}{\theta-it} + \frac{\alpha}{(\theta-it)^2} + \frac{2\alpha}{(\theta-it)^3} \right) + 2\lambda \left\{ \left(\frac{1}{(2\theta-it)} + \frac{\alpha}{(2\theta-it)^2} + \frac{2\alpha}{(2\theta-it)^3} \right) + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} \right\} \right\} \right)$$

respectively.

And hence show that Another Two-Parameter Sujatha distribution and Exponential distribution are particular cases of Transmuted another Two-Parameter Sujatha distribution.

PROOF: We begin with the well-known definition of the moment generating function given by

$$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x; \alpha, \theta, \lambda) dx$$

$$= \int_0^\infty e^{tx} \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1-\lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} dx$$

$$= \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \int_0^\infty e^{tx} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1-\lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} dx$$

$$M_X(t) = \left(\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left\{ (1-\lambda) \left(\frac{1}{\theta-t} + \frac{\alpha}{(\theta-t)^2} + \frac{2\alpha}{(\theta-t)^3} \right) + 2\lambda \left\{ \left(\frac{1}{(2\theta-t)} + \frac{\alpha}{(2\theta-t)^2} + \frac{2\alpha}{(2\theta-t)^3} \right) + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} \right\} \right\} \right)$$

(5.2.1)

Which is the m.g.f of Transmuted another Two-Parameter Sujatha distribution.

For $\lambda = 0$ in equation (5.3.1) we get

$$M_X(t) = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left(\frac{1}{\theta - t} + \frac{\alpha}{(\theta - t)^2} + \frac{2\alpha}{(\theta - t)^3} \right)$$

Which is m.g.f of *Another Two-Parameter Sujatha distribution* with parameters α, θ

For $\lambda = 0, \alpha = 0$ in equation (5.3.1) we get

$$M_X(t) = \frac{\theta}{(\theta - t)}$$

Which is the m.g.f of Exponential distribution

Also we know that $\varphi_X(t) = M_X(it)$

Therefore,

$$\varphi_X(t) = \left(\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left\{ \left((1 - \lambda) \left(\frac{1}{\theta - it} + \frac{\alpha}{(\theta - it)^2} + \frac{2\alpha}{(\theta - it)^3} \right) + 2\lambda \right) \left(\frac{1}{(2\theta - it)} + \frac{\alpha}{(2\theta - it)^2} + \frac{2\alpha}{(2\theta - it)^3} \right) + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} \right\} \left\{ \theta \left(\frac{2}{(2\theta - it)^3} + \frac{6\alpha}{(2\theta - it)^4} + \frac{24\alpha}{(2\theta - it)^5} \right) + (\alpha\theta + 2) \left(\frac{1}{(2\theta - it)^2} + \frac{2\alpha}{(2\theta - it)^3} + \frac{6\alpha}{(2\theta - it)^4} \right) \right\} \right) \quad (5.2.2)$$

Which is the characteristic function of Transmuted another Two-Parameter Sujatha distribution.

5.3 Mean deviation about mean and median of TATPSD

We have derived the expressions for mean deviation about mean and median of TATPSD in this section.

THEOREM 5.3.1: If X has the TATPSD(θ, α, λ), then the mean deviation about mean ($\delta_1(X)$) and mean deviation about median ($\delta_2(X)$) are given as:

$$\delta_1(X) = \left[\begin{array}{l} 2\mu \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\} \\ -2 \left\{ \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A + 2\lambda \left\{ B + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C + (\alpha\theta + 2)D) \right\} \right] \right\} \end{array} \right]$$

Where

$$A = \frac{1}{\theta^4} \{ \theta^2 + 2\theta\alpha + 6\alpha - e^{-\theta\mu} (\theta^2 + \mu\theta^3 + 2\alpha\theta + 2\mu\alpha\theta^2 + \mu^2\theta^3\alpha + 6\alpha\theta\mu + 6\alpha + 3\mu^2\theta^2\alpha + \mu^3\theta^3\alpha) \}$$

$$B = \left(\frac{1}{16\theta^4} \{ (4\theta^2 + 4\alpha\theta + 6\alpha - e^{-2\theta\mu} (4\theta^2 + 8\mu\theta^3 + 4\alpha\theta + 8\mu\alpha\theta^2 + 8\mu^2\theta^3\alpha + 12\alpha\theta\mu + 6\alpha + 12\mu^2\theta^2\alpha + 8\mu^3\theta^3\alpha)) \} \right)$$

$$C = \frac{1}{64\theta^5} \left\{ \begin{aligned} &24\theta^2 + 48\alpha\theta + 120\alpha - e^{-2\theta\mu} \\ &(24\theta^2 + 48\mu\theta^3 + 48\alpha\theta + 48\mu^2\theta^4 + 32\mu^3\theta^5 + 96\mu\theta^2\alpha + 96\mu^2\theta^3\alpha + 64\mu^3\theta^4\alpha) \\ &+ 32\mu^4\theta^5\alpha + 32\mu^5\theta^5\alpha + 240\mu\theta\alpha + 120\alpha + 240\theta^2\mu^2\alpha + 160\mu^3\theta^3\alpha + 80\mu^4\theta^4\alpha \end{aligned} \right\}$$

$$D = \frac{1}{32\theta^5} \left\{ \begin{aligned} &8\theta^2 + 12\alpha\theta + 24\alpha - e^{-2\theta\mu} \\ &(8\theta^2 + 16\mu\theta^3 + 12\alpha\theta + 16\mu^2\theta^4 + 24\mu\theta^2\alpha + 24\mu^2\theta^3\alpha + 16\mu^3\theta^4\alpha) \\ &+ 48\mu\theta\alpha + 24\alpha + 48\theta^2\mu^2\alpha + 32\mu^3\theta^3\alpha + 16\mu^4\theta^4\alpha \end{aligned} \right\}$$

And

$$\delta_2(X) = \left[\mu - 2 \left\{ \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A' + 2\lambda \left\{ B' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C' + (\alpha\theta + 2)D') \right\} \right] \right\} \right]$$

Where

$$A' = \left(\frac{1}{\theta^4} \{ \theta^2 + 2\theta\alpha + 6\alpha - e^{-\theta M} (\theta^2 + M\theta^3 + 2\alpha\theta + 2M\alpha\theta^2 + M^2\theta^3\alpha + 6\alpha\theta M + 6\alpha + 3M^2\theta^2\alpha + M^3\theta^3\alpha) \} \right)$$

$$B' = \left(\frac{1}{16\theta^4} \{ (4\theta^2 + 4\alpha\theta + 6\alpha - e^{-2\theta M} (4\theta^2 + 8M\theta^3 + 4\alpha\theta + 8M\alpha\theta^2 + 8M^2\theta^3\alpha + 12\alpha\theta M + 6\alpha + 12M^2\theta^2\alpha + 8M^3\theta^3\alpha)) \} \right)$$

$$C' = \left(\frac{1}{64\theta^5} \left\{ \begin{aligned} &24\theta^2 + 48\alpha\theta + 120\alpha - e^{-2\theta M} \\ &(24\theta^2 + 48M\theta^3 + 48\alpha\theta + 48M^2\theta^4 + 32M^3\theta^5 + 96M\theta^2\alpha + 96M^2\theta^3\alpha + 64M^3\theta^4\alpha) \\ &+ 32M^4\theta^5\alpha + 32M^5\theta^5\alpha + 240M\theta\alpha + 120\alpha + 240\theta^2M^2\alpha + 160M^3\theta^3\alpha + 80M^4\theta^4\alpha \end{aligned} \right\} \right)$$

$$D' = \frac{1}{32\theta^5} \left\{ \begin{aligned} &8\theta^2 + 12\alpha\theta + 24\alpha - e^{-2\theta M} \\ &(8\theta^2 + 16M\theta^3 + 12\alpha\theta + 16M^2\theta^4 + 24M\theta^2\alpha + 24M^2\theta^3\alpha + 16M^3\theta^4\alpha) \\ &+ 48M\theta\alpha + 24\alpha + 48\theta^2M^2\alpha + 32M^3\theta^3\alpha + 16M^4\theta^4\alpha \end{aligned} \right\}$$

respectively.

Proof: Mean deviation about mean and mean deviation about median are defined as

$$\delta_1(X) = \int_0^\infty |x - \mu| f_\tau(x, \alpha, \theta, \lambda) dx$$

$$\text{And } \delta_2(X) = \int_0^\infty |x - M| f_\tau(x, \alpha, \theta, \lambda) dx$$

respectively.

Where μ and M are mean and median respectively of random variable $X \sim \text{TATPSD}$. The measures $\delta_1(X)$ and $\delta_2(X)$ can be obtained by using the simplified relationships.

$$\begin{aligned} \delta_1(X) &= \int_0^\mu (\mu - x) f_\tau(x, \alpha, \theta, \lambda) dx + \int_\mu^\infty (x - \mu) f_\tau(x, \alpha, \theta, \lambda) dx \\ \delta_1(X) &= 2\mu F_\tau(\mu) - 2 \int_0^\mu x f_\tau(x, \alpha, \theta, \lambda) dx \end{aligned} \quad (5.3.1)$$

And

$$\begin{aligned} \delta_2(X) &= \int_0^M (M - x) f_\tau(x, \alpha, \theta, \lambda) dx + \int_M^\infty (x - M) f_\tau(x, \alpha, \theta, \lambda) dx \\ \delta_2(X) &= \mu - 2 \int_0^M x f_\tau(x, \alpha, \theta, \lambda) dx \end{aligned} \quad (5.3.2)$$

Where

$$f_\tau(x, \alpha, \theta, \lambda) = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\}$$

Now

$$\int_0^\mu x f_\tau(x, \alpha, \theta, \lambda) dx = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A + 2\lambda \left\{ B + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C + (\alpha\theta + 2)D) \right\} \right] \quad (5.3.3)$$

Where

$$A = \frac{1}{\theta^4} \{ \theta^2 + 2\theta\alpha + 6\alpha - e^{-\theta\mu}(\theta^2 + \mu\theta^3 + 2\alpha\theta + 2\mu\alpha\theta^2 + \mu^2\theta^3\alpha + 6\alpha\theta\mu + 6\alpha + 3\mu^2\theta^2\alpha + \mu^3\theta^3\alpha) \}$$

$$B = \frac{1}{16\theta^4} \{ 4\theta^2 + 4\alpha\theta + 6\alpha - e^{-2\theta\mu} (4\theta^2 + 8\mu\theta^3 + 4\alpha\theta + 8\mu\alpha\theta^2 + 8\mu^2\theta^3\alpha + 12\alpha\theta\mu + 6\alpha + 12\mu^2\theta^2\alpha + 8\mu^3\theta^3\alpha) \}$$

$$C = \left| \frac{1}{64\theta^5} \left\{ \begin{aligned} &24\theta^2 + 48\alpha\theta + 120\alpha - e^{-2\theta\mu} \\ &(24\theta^2 + 48\mu\theta^3 + 48\alpha\theta + 48\mu^2\theta^4 + 32\mu^3\theta^5 + 96\mu\theta^2\alpha + 96\mu^2\theta^3\alpha + 64\mu^3\theta^4\alpha \\ &+ 32\mu^4\theta^5\alpha + 32\mu^5\theta^5\alpha + 240\mu\theta\alpha + 120\alpha + 240\theta^2\mu^2\alpha + 160\mu^3\theta^3\alpha + 80\mu^4\theta^4\alpha) \end{aligned} \right\} \right|$$

$$D = \frac{1}{32\theta^5} \left\{ \begin{aligned} &8\theta^2 + 12\alpha\theta + 24\alpha - e^{-2\theta\mu} \\ &(8\theta^2 + 16\mu\theta^3 + 12\alpha\theta + 16\mu^2\theta^4 + 24\mu\theta^2\alpha + 24\mu^2\theta^3\alpha + 16\mu^3\theta^4\alpha) \\ &+ 48\mu\theta\alpha + 24\alpha + 48\theta^2\mu^2\alpha + 32\mu^3\theta^3\alpha + 16\mu^4\theta^4\alpha \end{aligned} \right\}$$

And

$$\int_0^M x f_{\tau}(x, \alpha, \theta, \lambda) dx = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A' + 2\lambda \left\{ B' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C' + (\alpha\theta + 2)D') \right\} \right] \quad (5.3.4)$$

Where

$$A' = \frac{1}{\theta^4} \left\{ \theta^2 + 2\theta\alpha + 6\alpha - e^{-\theta M} \right. \\ \left. (\theta^2 + M\theta^3 + 2\alpha\theta + 2M\alpha\theta^2 + M^2\theta^3\alpha + 6\alpha\theta M + 6\alpha + 3M^2\theta^2\alpha + M^3\theta^3\alpha) \right\}$$

$$B' = \left(\frac{1}{16\theta^4} \left\{ 4\theta^2 + 4\alpha\theta + 6\alpha - e^{-2\theta M} \right. \right. \\ \left. \left. (4\theta^2 + 8M\theta^3 + 4\alpha\theta + 8M\alpha\theta^2 + 8M^2\theta^3\alpha + 12\alpha\theta M + 6\alpha + 12M^2\theta^2\alpha + 8M^3\theta^3\alpha) \right\} \right)$$

$$C' = \frac{1}{64\theta^5} \left\{ \begin{aligned} &24\theta^2 + 48\alpha\theta + 120\alpha - e^{-2\theta M} \\ &(24\theta^2 + 48M\theta^3 + 48\alpha\theta + 48M^2\theta^4 + 32M^3\theta^5 + 96M\theta^2\alpha + 96M^2\theta^3\alpha + 64M^3\theta^4\alpha \\ &+ 32M^4\theta^5\alpha + 32M^5\theta^5\alpha + 240M\theta\alpha + 120\alpha + 240\theta^2M^2\alpha + 160M^3\theta^3\alpha + 80M^4\theta^4\alpha) \end{aligned} \right\}$$

$$D' = \frac{1}{32\theta^5} \left\{ \begin{aligned} &8\theta^2 + 12\alpha\theta + 24\alpha - e^{-2\theta M} \\ &(8\theta^2 + 16M\theta^3 + 12\alpha\theta + 16M^2\theta^4 + 24M\theta^2\alpha + 24M^2\theta^3\alpha + 16M^3\theta^4\alpha) \\ &+ 48M\theta\alpha + 24\alpha + 48\theta^2M^2\alpha + 32M^3\theta^3\alpha + 16M^4\theta^4\alpha \end{aligned} \right\}$$

Using expressions (5.3.1), (5.3.2), (5.3.3) and (5.3.4) and expression for c.d.f (2.2)

we obtain mean deviation about mean ($\delta_1(X)$) and mean deviation about median

($\delta_2(X)$)

$$\delta_1(X) = \left[\begin{aligned} &2\mu \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\} \\ &- 2 \left\{ \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A + 2\lambda \left\{ B + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C + (\alpha\theta + 2)D) \right\} \right] \right\} \end{aligned} \right]$$

&

$$\delta_2(X) = \left[\mu - 2 \left\{ \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A' + 2\lambda \left\{ B' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C' + (\alpha\theta + 2)D') \right\} \right] \right\} \right]$$

6. BONFERRONI AND LORENZ CURVES AND INDICES OF TATPSD

The Bonferroni curve($B(p)$), Lorenz curve($L(p)$), Bonferroni index (B) and Gini index (G) have find applicability in fields of economics, demography, reliability, life testing and medical sciences. The Bonferroni and Lorenz curves are defined as

$$B(p) = \frac{1}{p\mu} \int_0^q x f_{\tau}(x, \alpha, \theta, \lambda) dx \quad (6.1)$$

$$L(p) = \frac{1}{\mu} \int_0^q x f_{\tau}(x, \alpha, \theta, \lambda) dx \quad (6.2)$$

Where $\mu = E(x)$ is the mean of TATPSD and $q = F^{-1}(p)$.

The Bonferroni and Gini indices are defined as

$$B = 1 - \int_0^1 B(p) dp \quad (6.3)$$

$$G = 1 - 2 \int_0^1 L(p) dp \quad (6.4)$$

Using the p.d.f (2.4) of TATPSD we get

$$\int_0^q x f_{\tau}(x, \alpha, \theta, \lambda) dx = \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A'' + 2\lambda \left\{ B'' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C'' + (\alpha\theta + 2)D'') \right\} \right] \quad (6.5)$$

Where

$$A'' = \frac{1}{\theta^4} \left\{ \theta^2 + 2\theta\alpha + 6\alpha - e^{-\theta q} \right. \\ \left. \left((\theta^2 + q\theta^3 + 2\alpha\theta + 2q\alpha\theta^2 + q^2\theta^3\alpha + 6\alpha\theta q + 6\alpha + 3q^2\theta^2\alpha + q^3\theta^3\alpha) \right) \right\}$$

$$B'' = \frac{1}{16\theta^4} \left\{ 4\theta^2 + 4\alpha\theta + 6\alpha - e^{-2\theta q} \right. \\ \left. \left((4\theta^2 + 8q\theta^3 + 4\alpha\theta + 8q\alpha\theta^2 + 8q^2\theta^3\alpha + 12\alpha\theta q + 6\alpha + 12q^2\theta^2\alpha + 8q^3\theta^3\alpha) \right) \right\}$$

$$C'' = \frac{1}{64\theta^5} \left\{ \begin{aligned} &24\theta^2 + 48\alpha\theta + 120\alpha - e^{-2\theta q} \\ &(24\theta^2 + 48q\theta^3 + 48\alpha\theta + 48q^2\theta^4 + 32q^3\theta^5 + 96q\theta^2\alpha + 96q^2\theta^3\alpha + 64q^3\theta^4\alpha) \\ &+ 32q^4\theta^5\alpha + 32q^5\theta^5\alpha + 240q\theta\alpha + 120\alpha + 240\theta^2q^2\alpha + 160q^3\theta^3\alpha + 80q^4\theta^4\alpha \end{aligned} \right\}$$

$$D'' = \frac{1}{32\theta^5} \left\{ \begin{aligned} &8\theta^2 + 12\alpha\theta + 24\alpha - e^{-2\theta q} \\ &(8\theta^2 + 16q\theta^3 + 12\alpha\theta + 16q^2\theta^4 + 24q\theta^2\alpha + 24q^2\theta^3\alpha + 16q^3\theta^4\alpha) \\ &+ 48q\theta\alpha + 24\alpha + 48\theta^2q^2\alpha + 32q^3\theta^3\alpha + 16q^4\theta^4\alpha \end{aligned} \right\}$$

Using equation (6.5) in (6.1) & (6.2) we get

$$B(p) = \frac{1}{p\mu} \left[\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A'' + 2\lambda \left\{ B'' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C'' + (\alpha\theta + 2)D'') \right\} \right] \right] \quad (6.6)$$

And

$$L(p) = \frac{1}{\mu} \left[\left[\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A'' + 2\lambda \left\{ B'' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C'' + (\alpha\theta + 2)D'') \right\} \right] \right] \right] \quad (6.7)$$

Using (6.6) & (6.7) in (6.3) & (6.4) we get

$$B = 1 - \frac{1}{\mu} \left[\left[\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A'' + 2\lambda \left\{ B'' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C'' + (\alpha\theta + 2)D'') \right\} \right] \right] \right] \quad (6.8)$$

$$L = 1 - \frac{2}{\mu} \left[\left[\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} \left[(1 - \lambda)A'' + 2\lambda \left\{ B'' + \frac{\alpha\theta}{(\theta^2 + \alpha\theta + 2\alpha)} (C'' + (\alpha\theta + 2)D'') \right\} \right] \right] \right] \quad (6.9)$$

7. ORDER STATISTICS OF TATPSD

Assuming $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$ to be the ordered statistics of the random sample $x_1, x_2, x_3, \dots, x_n$ obtained from the Transmuted another Two-Parameter Sujatha distribution with cumulative distribution function $F_\tau(x, \alpha, \theta, \lambda)$ and probability density function $f_\tau(x, \alpha, \theta, \lambda)$, then the probability density function of r^{th} order statistics $X_{(r)}$ is given by:

$$f_{(r)}(x, \alpha, \theta, \lambda) = \frac{n!}{(r-1)!(n-r)!} f(x, \alpha, \theta, \lambda) [F(x, \alpha, \theta, \lambda)]^{r-1} [1 - F(x, \alpha, \theta, \lambda)]^{n-r},$$

$r=1, 2, 3, \dots, n$

Using the equations (2.2) and (2.4), the probability density function of r^{th} order statistics of transmuted another Two-Parameter Sujatha distribution is given by:

$$f_{(r)}(x, \alpha, \theta, \lambda) = \left[\frac{n!}{(r-1)!(n-r)!} \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right] \left[\left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\} \right]^{r-1} \left[1 - \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\} \right]^{n-r} .$$

Then, the p.d.f of first order statistic $X_{(1)}$ of Transmuted another Two-Parameter Sujatha distribution is given by:

$$f_{(1)}(x, \alpha, \theta, \lambda) = \left[n \frac{\theta^2}{(\alpha\theta + 2)^2} \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right] \left[1 - \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\} \right]^{n-1} .$$

and the pdf of n^{th} order statistic $X_{(n)}$ of Transmuted another Two-Parameter Sujatha distribution is given as:

$$f_{(n)}(x, \alpha, \theta, \lambda) = \left[n \frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right] \left[\left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \left\{ (1 + \lambda) - \lambda \left\{ 1 - \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\} \right]^{n-1} .$$

8. ESTIMATION OF PARAMETERS OF TATPSD

Assuming $x_1, x_2, x_3, \dots, x_n$ to be the random sample of size n drawn from Transmuted another Two-Parameter Sujatha distribution having density function given by (2.4), then the likelihood function of Transmuted another Two-Parameter Sujatha distribution is given as:

$$L(x|\alpha, \theta, \lambda) = \prod_{i=1}^n \left[\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)} (1 + \alpha x + \alpha x^2) e^{-\theta x} \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right]$$

The log likelihood function becomes:

$$\log L = \left(\begin{aligned} & \left\{ 3n \log \theta - n \log(\theta^2 + \alpha\theta + 2\alpha) - \theta \sum_{i=1}^n \{1 + \alpha x_i + \alpha x_i^2\} x_i + \sum_{i=1}^n \log \{ \} \right\} \\ & + \sum_{i=1}^n \log \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x_i} \right\} \end{aligned} \right) \tag{8.1}$$

Differentiating the log-likelihood function with respect to α , θ and λ . This is done by partially differentiate (8.1) with respect to θ, α and λ equating the result to zero, we obtain the following normal equations,

$$\frac{\partial \log L}{\partial \theta} = \left[\begin{aligned} & \left[\frac{3n}{\theta} - \frac{n(2\theta + \alpha)}{(\theta^2 + \alpha\theta + 2\alpha)} - \sum_{i=1}^n x_i \right. \\ & \left. \left\{ \frac{2\lambda e^{-\theta x_i} \left(\frac{(\theta^2 + \alpha\theta + 2\alpha)(2\alpha\theta x_i^2 + 2\theta\alpha^2 x_i + 2\alpha x_i) - (\alpha\theta x_i(\theta x_i + \alpha\theta + 2)(2\theta + \alpha))}{(\theta^2 + \alpha\theta + 2\alpha)^2} \right)}{-x_i \left(1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right)} \right\} \right] \\ & \left. + \sum_{i=1}^n \left\{ \frac{\left((1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x_i} \right)}{\left((1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x_i} \right)} \right\} \right] = 0 \tag{8.2} \end{aligned} \right.$$

$$\frac{\partial \log L}{\partial \alpha} = \left(\begin{aligned} & \left(-\frac{n(\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} + \sum_{i=1}^n \left(\frac{x_i + x_i^2}{1 + \alpha x_i + \alpha x_i^2} \right) \right) \\ & \left. + \sum_{i=1}^n \left\{ \frac{2\lambda e^{-\theta x_i} \{ (\theta^2 x_i^2 + 2\alpha\theta^2 x_i + 2\theta x_i)(\theta^2 + \alpha\theta + 2\alpha) - \alpha\theta x_i(\theta x_i + \alpha\theta + 2)(\theta + 2) \}}{(\theta^2 + \alpha\theta + 2\alpha)^2 \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x_i} \right\}} \right\} \right) = 0 \tag{8.3} \end{aligned} \right.$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^n \left[\frac{\left\{ 2 \left[1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x_i} - 1 \right\}}{\left\{ \left\{ (1 - \lambda) + 2\lambda \left[1 + \frac{\alpha\theta x_i(\theta x_i + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x_i} \right\} \right\}} \right] = 0 \tag{8.4}$$

MLEs of α, θ, λ cannot be obtained by solving above complex equations as these equations are not in closed form. So we solve above equations by using iteration method through R software.

9. MOTIVATION BEHIND PROPOSED MODEL

Addition of transmuted parameter has increased flexibility in terms of moments of distribution. So it was the flexibility of proposed model which is increased by transmuted parameter which motivated me to work on this model. Also proposed model has increasing as well as decreasing hazard rate as can be seen from graph 4, which is an interesting feature and finds greater applicability in real life.

10. APPLICATIONS OF TATPSD

We have analyzed two real life data sets to show that the transmuted another two-parameter Sujatha distribution can be a better model than another two-parameter Sujatha distribution, exponential distribution and Lindley Pareto distribution. We also tested the significance of transmuted parameter.

DATA SET 1: The data set given in table 1 is related to the burning velocity (cm/sec) of different chemical materials. The source of the data set related to the burning velocity of different chemical materials for the year 2005 is available on the website and has been used by Sajid and Riyaz (2014) [12].

TABLE 1: Burning velocity (cm/sec) of different chemical materials

68	61	64	55	51	68	44	82
60	89	61	54	166	66	50	87
48	42	58	46	67	46	46	44
48	56	47	54	47	80	38	108
46	40	44	312	41	31	40	41
40	56	45	43	46	46	46	46
52	58	82	71	48	39	41	

DATA SET 2: The data set given in table 2 represents the survival times (in months). This data is non-censored data for lung cancer patients obtained from Pena (2002). This data set was recently used by L. S. Diab & E. S. El-Atfy (2017) in

paper “A moment inequality for overall decreasing life class of life distributions with hypothesis testing applications” [13].

TABLE 2: Survival times (in months) for lung cancer patients.

0.99	1.28	1.77	1.97	2.17	2.63	2.66	2.76	2.79	2.86
2.99	3.06	3.15	3.45	3.71	3.75	3.81	4.11	4.27	4.34
4.40	4.63	4.73	4.93	4.93	5.03	5.16	5.17	5.49	5.68
5.72	5.85	5.98	8.15	8.62	8.48	8.61	9.46	9.53	10.05
10.15	10.94	10.94	11.24	11.63	12.26	12.65	12.78	13.18	13.47
13.96	14.88	15.05	15.31	16.13	16.46	17.45	17.61	18.20	18.37
19.06	20.70	22.54	23.36						

These data sets are used here only for illustrative purposes. The required numerical evaluations are carried out using R software version R 3.3.5. We have fitted another two-parameter Sujatha distribution, exponential distribution, Lindley Pareto distribution and transmuted another two-parameter Sujatha distribution to these data sets. The summary statistics of the data set 1 & 2 is displayed in table 3, MLEs of the parameters, model functions are displayed in table 4 for both the data sets and the corresponding log-likelihood values, LR statistic, AIC, AICC, BIC, HQIC, Kolmogorov statistic and p value are displayed in Table 5 & 6 for data sets 1 & 2 respectively.

TABLE 3: Summary statistic of data sets 1 & 2.

Data Set	No. of observations	Min.	First quartile	median	mean	Third quartile	Max.
Data Set 1	55	31.0	44.5	48.0	61.0	62.5	312.0
Data Set 2	64	0.990	3.795	5.915	8.710	12.880	23.360

TABLE 4: ML Estimates, Standard Error of Estimates in parenthesis, model function of proposed model and related models for data sets 1 & 2.

Data set	Distribution	ML estimates (standard errors)	Model Function
1	Transmuted Another Two-Parameter Sujatha Distribution (TATPSD)	$\hat{\theta} = 0.06406471$ $\hat{\alpha} = 0.29756831$ (0.63519694) $\hat{\lambda} = -1.00000000$	$\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)}(1 + \alpha x + \alpha x^2)e^{-\theta x}$ $\left\{ (1 - \lambda) \right.$ $\left. + 2\lambda \left\{ \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\}$
	Another Two-Parameter Sujatha Distribution (TATPSD)	$\hat{\theta} = 0.04879202$ (0.00379731) $\hat{\alpha} = 7.88106111$ (75.65217391)	$\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)}(1 + \alpha x + \alpha x^2)e^{-\theta x}$
	Exponential Distribution (ED)	$\hat{\theta} = 61.000000$ (8.225236)	$\frac{1}{\theta} e^{-\frac{x}{\theta}}$
	Lindley Pareto Distribution (LPD)	$\hat{\theta} = 0.007626414$ (0.003924553) $\hat{\alpha} = 0.916469935$ (0.470474649) $\hat{k} = 1.313641828$ (0.107580641)	$\frac{k\theta^2 e^{\theta} x^{2k-1}}{(\theta + 1)\alpha^{2k}} e^{-\theta(\frac{x}{\alpha})^k}$
2	Transmuted Another Two-Parameter Sujatha Distribution (TATPSD)	$\hat{\theta} = 0.28884058$ (0.01689813) $\hat{\alpha} = 0.02223941$ $\hat{\lambda} = -1.00000000$	$\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)}(1 + \alpha x + \alpha x^2)e^{-\theta x}$ $\left\{ (1 - \lambda) \right.$ $\left. + 2\lambda \left\{ \left[1 + \frac{\alpha\theta x(\theta x + \alpha\theta + 2)}{(\theta^2 + \alpha\theta + 2\alpha)} \right] e^{-\theta x} \right\} \right\}$
	Another Two-Parameter Sujatha Distribution (ATPSD)	$\hat{\theta} = 0.3254563$ $\hat{\alpha} = 3.2962191$	$\frac{\theta^3}{(\theta^2 + \alpha\theta + 2\alpha)}(1 + \alpha x + \alpha x^2)e^{-\theta x}$

Exponential Distribution (ED)	$\hat{\theta} = 8.71000$ (1.08875)	$\frac{1}{\theta} e^{-\frac{x}{\theta}}$
Lindley Pareto Distribution (LPD)	$\hat{\theta} = 0.02029546$ (0.02490351) $\hat{\alpha} = 0.08779542$ (0.11279414) $\hat{k} = 0.99870938$ (0.09371982)	$\frac{k\theta^2 e^{\theta} x^{2k-1}}{(\theta + 1)\alpha^{2k}} e^{-\theta(\frac{x}{\alpha})^k}$

TABLE 5: Model comparison and Likelihood ratio statistic of related models and proposed model for data set 1.

Distribution	$-\log L$	AIC	BIC	AICC	HQIC	K-S (D)	P Value	Likelihood Ratio
TATPSD	251.695	509.390	515.41	509.861	511.719	0.1888	0.0395	14.48
ATPSD	258.935	521.871	525.88	522.102	523.424	0.2246	0.0077	
ED	281.098	564.196	566.20	564.271	564.972	0.4454	6.628e-10	
LPD	260.174	526.348	532.37	526.819	528.677	0.2523	0.00181	

TABLE 6: Model comparison and Likelihood ratio statistic of related models and proposed model for data set 2.

Distribution	$-\log L$	AIC	BIC	AICC	HQIC	K-S (D)	P Value	Likelihood Ratio
TATPSD	193.137	392.274	398.75	392.674	394.826	0.1156	0.358	5.936
ATPSD	196.105	396.210	400.52	396.407	397.911	0.2542	0.0005	
ED	202.526	407.052	409.21	407.116	407.902	0.1825	0.0281	
LPD	194.481	394.963	401.43	395.363	397.514	0.1162	0.352	

For testing the significance of transmuted parameter λ of our proposed model and for checking superiority of Transmuted Another Two-Parameter Sujatha Distribution over Another Two-Parameter Sujatha Distribution, Lindley Pareto distribution and Exponential Distribution for given data sets 1 & 2 we computed likelihood ratio (LR) statistic. For testing $H_0: \lambda = 0$ versus $H_1: \lambda \neq 0$ the LR statistic for testing H_0 is $\omega_1 = 2\{L(\hat{\theta}) - L(\hat{\theta}_0)\} = 14.48$ for data set 1 & $\omega_2 = 2\{L(\hat{\theta}) - L(\hat{\theta}_0)\} = 5.936$ for data set 2 where $\hat{\theta}$ and $\hat{\theta}_0$ are MLEs under H_1 and H_0 . LR statistic $\omega \sim (\chi_{(1)}^2(\alpha = 0.05) = 3.841)$ as $n \rightarrow \infty$, where 1 = degrees of freedom is the difference in dimensionality. From table 5, $\omega_1 = 14.48 > 3.841$ & $\omega_2 = 5.936 > 3.841$ at 5% level of significance for both the data sets, so we reject H_0 and conclude that transmuted parameter λ plays statistically a significant role.

Further in order to compare the proposed model with related models, we consider the criteria like AIC (Akaike information criterion), AICC (corrected Akaike information criterion), BIC (Bayesian information criterion) and HQIC. The better distribution corresponds to lesser AIC, AICC, BIC and HQIC values.

$$\text{AIC} = 2k - 2\log L \qquad \text{AICC} = \text{AIC} + \frac{2k(k+1)}{n-k-1}$$

$$\text{BIC} = k \log n - 2\log L \qquad \text{HQIC} = 2k \log(\log(n)) + 2 \log L$$

where k is the number of parameters in the statistical model, n is the sample size and $-2\log L$ is the maximized value of the log-likelihood function under the considered model. From Table 5 & 6, it has been observed that the Transmuted Another Two-Parameter Sujatha distribution possesses the lesser AIC, AICC BIC and HQIC values as compared to Another Two-Parameter Sujatha distribution, Lindley Pareto distribution & Exponential distribution for data sets 1 & 2. Hence we can conclude that the Transmuted Another Two-Parameter Sujatha distribution leads to a better fit than Another Two-Parameter Sujatha distribution and exponential distribution for data sets 1 & 2

For testing the goodness of fit of our proposed model Transmuted Another Two Parameter Sujatha distribution over its related models Another Two-Parameter Sujatha distribution, Lindley Pareto distribution & Exponential distribution to the

two data sets we computed Kolmogorov statistic and p value. The better model possesses lesser Kolmogorov statistic value and higher p value. It can be seen from tables 5 & 6 that Transmuted Another Two-Parameter Sujatha distribution possesses lesser Kolmogorov statistic value and higher p value for both the data sets as compared to Another Two-Parameter Sujatha distribution, Lindley Pareto distribution & Exponential distribution.

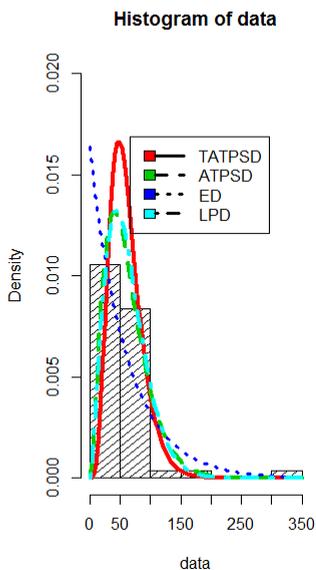


Figure 5 Curve fitting of data set 1

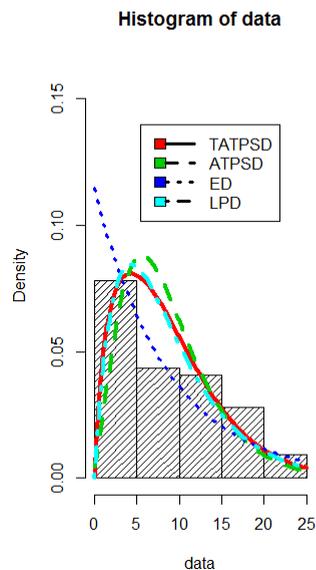


Figure 6 curve fitting of data set 2

11. CONCLUSION

We developed Transmuted another Two-Parameter Sujatha Distribution by using Quadratic Rank Transmutation Map technique. The various important properties of this distribution have been obtained. From the p.d.f plots (graphs 2(a) & 2(b)) it has been observed that our proposed model is positively skewed. Various reliability measures have been obtained for the proposed model. We fitted our model and its related models to two real life data sets and found that our proposed model gives better results for both the data sets than its related models as Kolmogorov statistic value, AIC, BIC, AICC, HQIC are lesser for proposed model on the both data sets than base model and exponential distribution. The significance of

transmuted parameter has been tested and it has been concluded that transmuted parameter plays a significant role. Researchers in future can generalize the proposed model and proposed model finds applicability in life testing.

12. BENEFITS OF THE PROPOSED MODEL

The proposed model can be used in variety of real life situations. As can be seen from application part of the paper, proposed model can be used for modeling of survival time's data. Also statistical significance of transmuted parameter shows the importance of proposed model.

REFERENCES

- [1] W. Shaw and I. Buckley, *The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions and a Skew-Kurtotic- Normal Distribution from a Rank Transformation Map*, arX. prer. arX. (2007), 0901.0434.
- [2] G. R. Aryal and C. P. Tsokos, *On the Transmuted Extreme Value Distribution with Applications*, Nonl. Anal. Th. Meth. Appl. 71 (2009), no. 12, 1401-1407.
- [3] A. Hassan, S. A. Wani and B. A. Para, *On three Parameter Weighted Quasi Lindley Distribution: Properties and Applications*, Int. Jou. Sci. Res. Math. Stat. Sci, 5 (2018), no. 5, 210-224.
- [4] A. Hassan, S. A. Wani, S. Shafi and B. A. Sheikh, *Lindley-Quasi Xgamma Distribution: Properties and Applications*, Pak. Jou. Stat. 36 (2020), no. 1, 73-89.
- [5] F. Merovci, *Transmuted Rayleigh Distribution*, Aust. Jou. Stat. 42 (2013), no. 1, 21-31.
- [6] B. A. Para and T. R. Jan, *A New Generalized Version of Log-logistic Distribution with Applications in Medical Sciences and other Applied Fields*, Int. Jou. Mod. Sim. Sci. Comp. 9 (2018), no. 5.
- [7] M. A. Haq, *On Transmuted Exponentiated Inverse Rayleigh Distribution*, Jou. Stat. Appl. Prob. 5 (2016), no. 2, 337-343.
- [8] A. Hassan, S. A. Wani and S. Shafi, *Poisson Pranav Distribution and its Applications*, Pak. Jou. Stat. 36 (2020), no. 1, 57-72.
- [9] T. Mussie and R. Shanker, *Another Two-Parameter Sujatha Distribution with Properties and Applications*, Jou. Math. Sci. Mod. 2 (2019), no. 1, 1-13.
- [10] M. E., Ghitany, B. Atieh and S. Nadarajah, *Lindley Distribution and Its Applications*, Math. Comp. Simul. 78 (2008), no. 4, 493-506.
- [11] S. A. Lone, A. Rahman and A. Islam, *Step Stress Partially Accelerated Life Testing Plan for Competing Risk using Adaptive Type-I Progressive Hybrid Censoring*, Pak. Jou. Stat. 33 (2017), no. 4, 237-248.
- [12] S. Ali and M. Riaz, *On the generalized process capability under simple and mixture models*, Jou. Appl. Math. 41 (2014), no. 4, 832-852.
- [13] L. S. Diab and E. S. El-Atfy, *A Moment Inequality for Overall Decreasing Life Class of Life Distributions with Hypothesis Testing Applications*. Int. Jou. Math. Stat. Inv. 5 (2017), no. 1, 62-71.

Sameer Ahmad Wani

Department of Statistics, Government Degree College Sopore, India

Email: wanisameer199@gmail.com (Corresponding Author)

Showkat Ahmad Dar

Department of Statistics, University of Kashmir, Srinagar, 190006, India

Email: darshowkat2429@gmail.com

Corresponding Author: Sameer Ahmad Wani