Computational Journal of Mathematical and Statistical Sciences
2(1), 80–105
DOI:10.21608/cjmss.2023.190993.1003
https://cjmss.journals.ekb.eg/

Research article

Chen Burr-Hatke Exponential Distribution: Properties, Regressions and Biomedical Applications

Lea Anzagra[1], **Abdul Ghaniyyu Abubakari**[2*], **Suleman Nasiru**[2]

[1] Department of Applied Statistics, Faculty of Social Science and Arts, S. D. Dombo University of Business and Integrated Development Studies, Wa, Ghana.

[2] Department of Statistics and Actuarial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana.

* **Correspondence:** aabubakari@cktutas.edu.gh

Abstract: Accurately modeling lifetime data is very important for appropriate decision making in health and biomedical fields. This usually requires the use of distributions. However, no single distribution can model all types of data. Hence, the development of distributions with appropriate usefulness is very important for modeling purposes. In this study, a new lifetime distribution, known as Chen Burr-Hatke exponential distribution is proposed. The objective of the study is to obtain a new lifetime distribution which can serve as an alternative distribution to modeling lifetime data. Also, such a distribution can be used to provide inferences via regression models. Plots of the density function of the new distribution show that the distribution can exhibit increasing, decreasing, right-skewed and left-skewed shapes. Also, plots of the hazard rate function show that the distribution can exhibit increasing, decreasing, and upside down bathtub shapes. Statistical properties, such as the quantile function, moments, order statistics and inequality measures, are derived. Several estimation methods are used to estimate the parameters of the distribution. Using Monte Carlo simulations, the estimators were all consistent. However, maximum likelihood estimation method was observed to better estimate the parameters of the distribution. Two regression models based on the distribution are established. The usefulness of the distribution and its regression models are demonstrated using real lifetime datasets. The results show that the models can provide a good fit to lifetime data, and hence can serve as alternative models to fitting such data.

Keywords: Chen-G; Burr-Hatke exponential distribution; regression; link function; lifetime data
Mathematics Subject Classification: 62J02, 60E05, 92C55
Received: 31 January 2023; **Revised:** 14 February 2023; **Accepted:** 20 February 2023; **Published:** 23 February 2023.

1. Introduction

In biomedical sciences, the ability to appropriately describe variables related to the occurrence of diseases and other health conditions, is very essential for proper treatment and development of response systems. An inappropriate modeling of the data will result in wrong inference, which may have severe consequences, especially for disease control. Parametric distributions are useful in modeling data from these fields. Also, these distributions could be extended to model the effect of exogenous variables on a response variable in the form of regression models. Recently, several new distributions and regression models to provide inferences on these distributions have been developed for modeling health and biomedical data, among other fields. Some distributions and classes of distributions developed include exponentiated Burr XII Poisson distribution (da Silva et al. [1]), Weibull Burr XII (WBXII) distribution (Afify et al. [2]), odd log-logistic Topp–Leone G family of distributions (Alizadeh et al. [3]), Burr-Hatke exponential (BHE) distribution (Abouelmagd, [4]; Yadav et al. [5]), odd generalized gamma-G family of distributions (Nasir et al. [6]), Chen-G family of distributions (Anzagra et al. [7]), inverse-power Burr-Hatke distribution (Afify et al. [8]), harmonic mixture Weibull-G family of distributions (Zamanah et al. [9]), harmonic mixture G family of distributions (Kharazmi et al. [10]) and odd log-logistic Weibull-G family of distributions (Rasekhi et al. [11]).

In this study, the one parameter BHE distribution is modified using the Chen-G family of distributions proposed by Anzagra et al. [7]. The new distribution is known as the Chen Burr-Hatke exponential (CBHE) distribution. The BHE distribution, though has some usefulness, can only model datasets which exhibit decreasing or are right-skewed shaped (Yadav et al. [5]). The motivation for this study is in two folds. Firstly, this study seeks to improve upon the BHE distribution, making it more flexible for modeling various types of data that exhibit different characteristics. Secondly, the study seeks to develop regression models with different structures and link functions. This will enable researchers model the effect of exogenous variables on a response variable which follows the CBHE distribution. Thus, the objective of this study is to develop a more flexible alternative lifetime distribution, capable of modeling lifetime data and also provide inferences via regression models.

The remaining article is organized as follows: Section 2 presents the CBHE distribution and the expansion of its density function. Some statistical properties of the CBHE distribution are presented in Section 3. Section 4 presents methods for estimating the CBHE parameters. Monte Carlo simulations of the estimators are also presented in the section. The empirical applications of the CBHE distribution are presented in Section 5 using real datasets. Regression models with the response variable following the CBHE distribution are developed in Section 6. An empirical application of the regression models is also presented in the section. The conclusion of the study is presented in Section 7.

2. Chen Burr-Hatke Exponential Distribution

A random variable X is said to follow the BHE distribution if its cumulative distribution function (CDF) is given by

$$G(x) = 1 - \frac{e^{-\varsigma x}}{1 + \varsigma x}, x > 0, \varsigma > 0. \tag{2.1}$$

Also, the CDF of the Chen-G family of distributions (Anzagra et al. [7]) is given as

$$F(x) = P\left[1 - e^{\upsilon\left(1-e^{G(x)^{\xi}}\right)}\right], x > 0, \upsilon > 0, \xi > 0, \tag{2.2}$$

where $P = \left(1 - e^{\upsilon(1-e)}\right)^{-1}$ and $G(x)$ is the baseline distribution. Substituting the CDF of the BHE distribution given by equation (2.1) into the CDF of the Chen-G family of distributions given by equation (2.2), gives the CDF of the CBHE distribution as

$$F(x) = P\left[1 - \exp\left\{\upsilon\left[1 - \exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}\right], x > 0, \upsilon > 0, \xi > 0, \varsigma > 0, \tag{2.3}$$

where $P = \left(1 - e^{\upsilon(1-e)}\right)^{-1}$. Differentiating the CDF in equation (2.3) gives the probability density function (PDF) of the CBHE distribution as

$$f(x) = P\upsilon\varsigma\xi\frac{2+\varsigma x}{(1+\varsigma x)^2}\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi-1} e^{-\varsigma x}\exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\exp\left\{\upsilon\left[1 - \exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}, x > 0. \tag{2.4}$$

The survival and hazard rate functions of the CBHE distribution are, respectively, given by

$$s(x) = 1 - P\left[1 - \exp\left\{\upsilon\left[1 - \exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}\right], x > 0, \tag{2.5}$$

and

$$h(x) = \frac{P\upsilon\varsigma\xi\frac{2+\varsigma x}{(1+\varsigma x)^2}\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi-1} e^{-\varsigma x}\exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\exp\left\{\upsilon\left[1 - \exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}}{1 - P\left[1 - \exp\left\{\upsilon\left[1 - \exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}\right]}, x > 0. \tag{2.6}$$

Plots of the PDF of the CBHE distribution for different combinations of the parameter values are displayed in Figure 1. The plots exhibit varying shapes including left-skewed, right-skewed and reversed J shapes. This indicates the flexibility of the CBHE distribution.

Figure 2 shows the plots of the hazard rate function of the CBHE distribution. The plots exhibit decreasing, increasing and upside-down bathtub failure shapes. Again, this indicates the flexibility of the distribution.

The mixture representation of the PDF of the CBHE distribution is obtained. The mixture representation is useful for the derivation of some properties of the distribution, especially properties involving the integration of the PDF. The mixture presentation is obtained as follows. Using Taylor series expansion, given as $e^z = \sum_{i=0}^{\infty} \frac{z^i}{i!}$, and binomial expansion, given as $(1-a)^i = \sum_{j=0}^{i} \binom{i}{j} a^j (-1)^j, |a| < 1$, gives

$$\exp\left\{\upsilon\left[1 - \exp\left\{\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\} = \sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}\frac{\upsilon^i}{i!}(-1)^{i+j}\exp\left\{(i-j)\left(1 - \frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}. \tag{2.7}$$

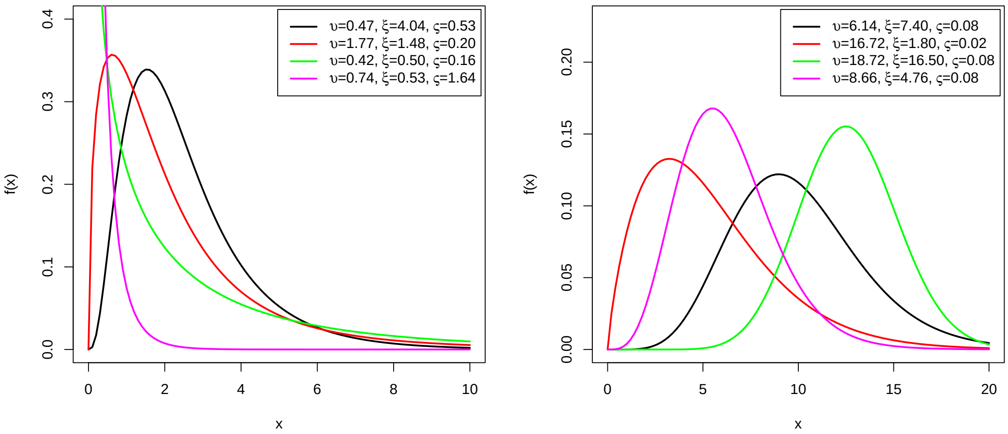

Figure 1. Plots of the PDF of CBHE Distribution

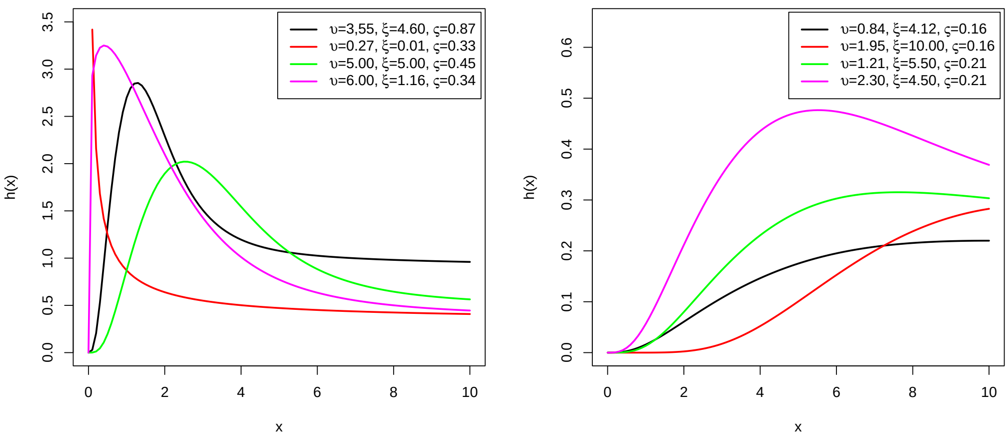

Figure 2. Plots of Hazard Rate Function of CBHE Distribution

Substituting equation (2.7) into the PDF in equation (2.4) gives the PDF as

$$f(x) = P\upsilon\varsigma\xi\frac{2+\varsigma x}{(1+\varsigma x)^2}\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi-1} e^{-\varsigma x}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}\frac{\upsilon^i}{i!}(-1)^{i+j}\exp\left\{(i-j+1)\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}. \tag{2.8}$$

Again, using Taylor series expansion gives equation (2.8) as

$$f(x) = P\upsilon\varsigma\xi\frac{2+\varsigma x}{(1+\varsigma x)^2}e^{-\varsigma x}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\binom{i}{j}\frac{\upsilon^i}{i!}(-1)^{i+j}(i-j+1)^k\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{-(1-\xi(k+1))}.$$

Furthermore, using binomial expansion gives

$$f(x) = P\upsilon\varsigma\xi(2+\varsigma x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\binom{i}{j}\binom{m-\xi(k+1)}{m}\frac{\upsilon^i}{i!}(-1)^{i+j}(i-j+1)^k\frac{e^{-(m+1)\varsigma x}}{(1+\varsigma x)^{m+2}}. \tag{2.9}$$

Using the Taylor series expansion defined as $z^{-\alpha} = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\alpha^{(n)}(z-1)^n$, where $\alpha^{(n)} = \alpha(\alpha+1)(\alpha+2)\dots(\alpha+n-1)$ is the rising factorial, gives the PDF in equation (2.9) as

$$f(x) = P\varsigma\xi(2+\varsigma x)\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\binom{i}{j}\binom{m-\xi(k+1)}{m}\frac{\upsilon^{i+1}}{i!n!}(-1)^{i+j+n}(i-j+1)^k(m+2)^{(n)}e^{-(m+1)\varsigma x}(\varsigma x)^n.$$

Thus, the PDF of the CBHE distribution can be written as

$$f(x) = (2+\varsigma x)\sum_{i=0}^{\infty}\Upsilon_{jkmn}x^n e^{-(m+1)\varsigma x}, \tag{2.10}$$

where $\Upsilon_{jkmn} = P\xi\sum_{j=0}^{i}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\binom{i}{j}\binom{m-\xi(k+1)}{m}\frac{\upsilon^{i+1}\varsigma^{n+1}}{i!n!}(-1)^{i+j+n}(i-j+1)^k(m+2)^{(n)}$.

3. Statistical properties

The statistical properties of the CBHE distribution are presented in this section. These include; the quantile function, ordinary moments, incomplete moments, moments generating function and order statistics.

3.1. *Quantile function*

The quantile function of a distribution is very useful for the characterization of the distribution, especially for generating random numbers. The quantile function is obtained as the inverse of the CDF of the distribution. Hence, the quantile function of the CBHE distribution is obtained by equating the CDF of the distribution in equation (2.3) to u and making x the subject of the equation as follows:

$$u = P\left[1-\exp\left\{\upsilon\left[1-\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}\right].$$

After some algebraic manipulations, we obtain

$$(1+\varsigma x)\, e^{1+\varsigma x} = e\left[1-\left[\log\left[1-\frac{1}{\upsilon}\log\left(1-\frac{u}{P}\right)\right]\right]^{\frac{1}{\xi}}\right]^{-1}.$$

Introducing the Lambert function, defined as $W(xe^x) = x$, yields

$$W\left((1+\varsigma x)\, e^{1+\varsigma x}\right) = W\left(e\left[1-\left[\log\left[1-\frac{1}{\upsilon}\log\left(1-\frac{u}{P}\right)\right]\right]^{\frac{1}{\xi}}\right]^{-1}\right).$$

Making x the subject and letting $Q(u) = x$, gives the quantile function of the CBHE distribution as

$$Q(u) = \frac{1}{\varsigma}\left[W\left(e\left[1-\left[\log\left[1-\frac{1}{\upsilon}\log\left(1-\frac{u}{P}\right)\right]\right]^{\frac{1}{\xi}}\right]^{-1}\right)-1\right], 0 \le u \le 1. \tag{3.1}$$

3.2. *Moments and Moments Generating Function*

The study of important statistical measures, such as measures of central tendencies, skewness and kurtosis among others, of a distribution require the use of moments of the distribution. The ordinary and incomplete moments, and moments generating function of the CBHE distribution are presented in this subsection.

3.2.1. Ordinary Moment

The r^{th} ordinary moment of the CBHE random variable is defined as $E(X^r) = \mu_r' = \int_0^\infty x^r f(x)dx, r = 1, 2, ...$ Substituting the PDF of the CBHE distribution in equation (2.10) into the definition gives

$$\begin{aligned}\mu_r' &= \sum_{i=0}^{\infty} \Upsilon_{jkmn} \int_0^\infty x^{r+n}\,(2+\varsigma x)\, e^{-(m+1)\varsigma x} dx,\\ &= \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{2\int_0^\infty x^{r+n} e^{-(m+1)\varsigma x} dx + \varsigma \int_0^\infty x^{r+n+1} e^{-(m+1)\varsigma x} dx\right\}.\end{aligned} \tag{3.2}$$

Let $y = (m+1)\,\varsigma x$. This implies that $x = \frac{y}{y(m+1)\varsigma}$ and $dx = \frac{dy}{y(m+1)\varsigma}$. Also, as $x \to 0, y \to 0$ and as $x \to \infty, y \to \infty$. Substituting these into equation (3.2) and simplifying gives

$$\mu_r' = \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{\frac{2}{[(m+1)\,\varsigma]^{r+n+1}} \int_0^\infty y^{r+n} e^{-y} dy + \frac{\varsigma}{[(m+1)\,\varsigma]^{r+n+2}} \int_0^\infty y^{r+n+1} e^{-y} dy\right\}.$$

Using the gamma function defined as $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$, the r^{th} ordinary moment of the CBHE distribution is obtained as

$$\mu_r' = \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{\frac{2\Gamma(r+n+1)}{[(m+1)\,\varsigma]^{r+n+1}} + \frac{\varsigma\Gamma(r+n+2)}{[(m+1)\,\varsigma]^{r+n+2}}\right\}, r = 1, 2, ... \tag{3.3}$$

The mean of the CBHE distribution is obtained by letting $r = 1$ in equation (3.3). This gives

$$\mu = \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ \frac{2\Gamma(n+2)}{[(m+1)\varsigma]^{n+2}} + \frac{\varsigma\Gamma(n+3)}{[(m+1)\varsigma]^{n+3}} \right\}. \tag{3.4}$$

Central moments of the CBHE distribution, defined as $E[(X-\mu)^r], r = 1, 2, \ldots$, can be obtained using the ordinary moments. For $r = 2, 3$ and 4, the central moments are given as $E\left[(X-\mu)^2\right] = \sigma^2 = \mu_2' - \mu^2$, $E\left[(X-\mu)^3\right] = \mu_3' - 3\mu_2'\mu + 2\mu^3$ and $E\left[(X-\mu)^4\right] = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4$. Therefore, the coefficients of skewness and kurtosis of the distribution can be obtained, respectively, as

$$CS = \frac{E\left[(X-\mu)^3\right]}{\sigma^3} \quad \text{and} \quad CK = \frac{E\left[(X-\mu)^4\right]}{\sigma^4}.$$

Table 1 shows the first five ordinary moments, standard deviation (SD), coefficient of variation (CV), CS and CK of the CBHE distribution for the following selected parameter sets: $I = (\upsilon = 0.1, \xi = 30.5, \varsigma = 200)$, $II = (\upsilon = 0.01, \xi = 38.7, \varsigma = 80.1)$, $III = (\upsilon = 2.0, \xi = 1.1, \varsigma = 1.4)$ and $IV = (\upsilon = 0.2, \xi = 0.4, \varsigma = 1.3)$. It can be observed that the distribution can exhibit both right and left skewness, and various degrees of kurtosis.

Table 1. Moments, SD, CV, CS and CK of CBHE Distribution

μ_r'	I	II	III	IV
μ_1'	0.014775	0.039795	0.234099	0.284516
μ_2'	2.47×10^{-4}	0.001773	0.139608	0.267893
μ_3'	4.40×10^{-6}	8.87×10^{-5}	0.165835	0.441450
μ_4'	8.73×10^{-8}	4.88×10^{-6}	0.323470	1.052015
μ_5'	2.21×10^{-9}	3.12×10^{-7}	0.889695	3.295082
SD	0.005389	0.013754	0.291215	0.432369
CV	0.364739	0.345614	1.243982	1.519664
CS	-0.695284	1.187088	3.783779	3.202499
CK	9.564007	2.271058	28.514464	18.887401

3.2.2. Incomplete Moments

Incomplete moments are quite useful, especially in deriving other quantities such as inequality measures and mean residual life of a distribution. The r^{th} incomplete moment of a CBHE random variable is defined as $m_r(x) = \int_0^x y^r f(y)dy, r = 1, 2, \ldots$. Substituting the PDF of the CBHE distribution in equation (2.10) into the definition gives

$$\begin{aligned} m_r(x) &= \sum_{i=0}^{\infty} \Upsilon_{jkmn} \int_0^x y^{r+n} (2 + \varsigma y) e^{-(m+1)\varsigma y} dy, \\ &= \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ 2 \int_0^x y^{r+n} e^{-(m+1)\varsigma y} dx + \varsigma \int_0^x y^{r+n+1} e^{-(m+1)\varsigma y} dy \right\}. \end{aligned} \tag{3.5}$$

Let $u = (m+1)\varsigma y$. This implies that $y = \frac{u}{(m+1)\varsigma}$ and $dy = \frac{du}{(m+1)\varsigma}$. Also, as $y \to 0, u \to 0$ and as $y \to x, u \to (m+1)\varsigma x$. Substituting these into equation (3.5) gives

$$m_r(x) = \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ \frac{2}{[(m+1)\varsigma]^{r+n+1}} \int_0^{(m+1)\varsigma x} u^{r+n} e^{-u} du + \frac{\varsigma}{[(m+1)\varsigma]^{r+n+2}} \int_0^{(m+1)\varsigma x} u^{r+n+1} e^{-y} du \right\}.$$

Using the incomplete gamma function defined as $\gamma(a, y) = \int_0^y t^{a-1} e^{-t} dt$, the r^{th} incomplete moment of the CBHE distribution is obtained as

$$m_r(x) = \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ \frac{2\gamma(r+n+1, (m+1)\varsigma x)}{[(m+1)\varsigma]^{r+n+1}} + \frac{\varsigma\gamma(r+n+2, (m+1)\varsigma x)}{[(m+1)\varsigma]^{r+n+2}} \right\}, r = 1, 2, \ldots. \quad (3.6)$$

When $r = 1$ in equation (3.6), the first incomplete moment of the CBHE distribution is obtained as

$$m_1(x) = \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ \frac{2\gamma(n+2, (m+1)\varsigma x)}{[(m+1)\varsigma]^{n+2}} + \frac{\varsigma\gamma(n+3, (m+1)\varsigma x)}{[(m+1)\varsigma]^{n+3}} \right\}. \quad (3.7)$$

Lorenz and Bonferroni curves, which are used to study income inequality, can be obtained using the incomplete moment of a distribution. Lorenz and Bonferroni curves are defined, respectively, as

$$L(x) = \frac{1}{\mu} \int_0^x y f(y) dy = \frac{m_1(x)}{\mu} \quad \text{and} \quad B(x) = \frac{1}{\mu F(x)} \int_0^x y f(y) dy = \frac{m_1(x)}{\mu F(x)},$$

where μ and $m_1(x)$ are the first ordinary and first incomplete moments of the distribution. Thus, substituting the first incomplete moment in equation (3.7) into the definitions give the Lorenz and Bonferroni curves of the CBHE distribution, respectively, as

$$L(x) = \frac{1}{\mu} \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ \frac{2\gamma(n+2, (m+1)\varsigma x)}{[(m+1)\varsigma]^{n+2}} + \frac{\varsigma\gamma(n+3, (m+1)\varsigma x)}{[(m+1)\varsigma]^{n+3}} \right\}$$

and

$$B(x) = \frac{1}{\mu F(x)} \sum_{i=0}^{\infty} \Upsilon_{jkmn} \left\{ \frac{2\gamma(n+2, (m+1)\varsigma x)}{[(m+1)\varsigma]^{n+2}} + \frac{\varsigma\gamma(n+3, (m+1)\varsigma x)}{[(m+1)\varsigma]^{n+3}} \right\}.$$

Gastwirth [12] has shown that Lorenz curve can be defined as

$$L(u) = \int_0^u Q(t)\, dt, \quad u \in (0, 1),$$

where $Q(t)$ is the quantile function of the distribution. Similarly, the Bonferroni curve can be defined as

$$B(u) = \frac{L(u)}{u}, \quad u \in (0, 1).$$

Figure 3 shows the plots of Lorenz and Bonferroni curves of the CBHE distribution. Lorenz curves of a distribution are usually convex in shapes. This is demonstrated by the plots of the CBHE Lorenz curves. When $L(u) = u$, then the point of minimal inequality or equidistribution line is obtained.

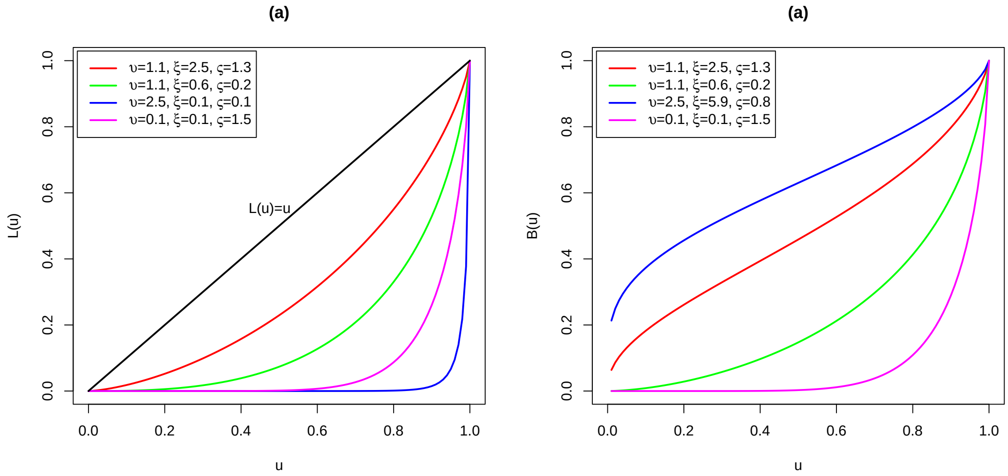

Figure 3. Plots of Lorenz (Left) and Bonferroni (Right) Curves

3.2.3. Moments Generating Function

The moment generating function (MGF) of a distribution, if it exists, is useful in deriving the moments of a distribution. MGF is defined as $M_X(t) = E(e^{tX})$. Using Taylor series expansion, MGF can be obtained as $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu'_r$, where μ'_r is the r^{th} ordinary moment of the distribution. Thus, substituting the r^{th} ordinary moment of the CBHE distribution in equation (3.3) into the definition gives the MGF of the CBHE distribution as

$$M_X(t) = \sum_{r=0}^{\infty}\sum_{i=0}^{\infty} \frac{t^r}{r!}\Upsilon_{jkmn}\left\{\frac{2\Gamma(r+n+1)}{[(m+1)\varsigma]^{r+n+1}} + \frac{\varsigma\Gamma(r+n+2)}{[(m+1)\varsigma]^{r+n+2}}\right\}. \tag{3.8}$$

3.3. *Order statistics*

The PDF for the p^{th} order statistic $x_{p:n}$, of an independent and identically distributed random sample $x_1, x_2, ..., x_n$ of size n, $f_{p:n}(x)$, is defined as

$$f_{p:n}(x) = \frac{n!}{(p-1)!(n-p)!}[F(x)]^{p-1}[1-F(x)]^{n-p} f(x), \quad p = 1, 2, ..., n.$$

Using binomial series expansion, the PDF can be written as

$$f_{p:n}(x) = \frac{n!}{(p-1)!(n-p)!}\sum_{l=0}^{n-p}(-1)^l \binom{n-p}{l}[F(x)]^{p+l-1} f(x). \tag{3.9}$$

From the CDF of the CBHE distribution in equation (2.3), we have

$$\begin{aligned}
[F(x)]^{p+l-1} &= P^{p+l-1}\left[1-\exp\left\{\upsilon\left[1-\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}\right]^{p+l-1}, \\
&= P^{p+l-1}\sum_{q=0}^{p+l-1}\binom{p+l-1}{q}(-1)^q \exp\left\{q\upsilon\left[1-\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}. \quad (3.10)
\end{aligned}$$

Substituting equation (3.10) and the PDF of the CBHE distribution in equation (2.4) into equation (3.9) gives the PDF of the p^{th} order statistic of the CBHE distribution as

$$f_{p:n}(x) = \frac{n!\upsilon\varsigma\xi\,(2+\varsigma x)\,e^{-\varsigma x}}{(p-1)!(n-p)!\,(1+\varsigma x)^2}\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi-1}\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}$$
$$\times\sum_{l=0}^{n-p}\sum_{q=0}^{p+l-1}\binom{n-p}{l}\binom{p+l-1}{q}(-1)^{l+q}P^{p+l}\exp\left\{(q+1)\,\upsilon\left[1-\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}.$$

Also, the PDF of the first and n^{th} order statistics of a distribution are defined, respectively, as

$$f_{1:n}(x) = n\sum_{l=0}^{n-1}\binom{n-1}{l}(-1)^j\,[F(x)]^j\,f(x) \quad\text{and}\quad f_{n:n}(x) = n\,[F(x)]^{n-1}\,f(x).$$

For the CBHE distribution, the PDF of the first and n^{th} order statistics are given, respectively, as

$$f_{1:n}(x) = \frac{n\upsilon\varsigma\xi\,(2+\varsigma x)\,e^{-\varsigma x}}{(1+\varsigma x)^2}\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi-1}\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}$$
$$\times\sum_{l=0}^{n-1}\sum_{q=0}^{j}\binom{n-1}{l}\binom{l}{q}(-1)^{l+q}P^{l+l}\exp\left\{(q+1)\,\upsilon\left[1-\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}$$

and

$$f_{n:n}(x) = \frac{n\upsilon\varsigma\xi\,(2+\varsigma x)\,e^{-\varsigma x}}{(1+\varsigma x)^2}\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi-1}\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}$$
$$\times\sum_{q=0}^{n-1}\binom{n-1}{q}(-1)^{q}P^{n}\exp\left\{(q+1)\,\upsilon\left[1-\exp\left\{\left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)^{\xi}\right\}\right]\right\}.$$

Figure 4 shows the minimum-maximum (min-max) plots of the order statistics of the CBHE distribution. The plots are based on the expectation of the first and n^{th} order statistics. The min-max plots gives information, such as skewness and kurtosis, of a distribution. From the plots, it can be observed that the CBHE distribution can exhibit symmetric, right-skewed and left-skewed shapes.

4. Parameter Estimation

The parameters of the CBHE distribution are estimated in this section using maximum likelihood, maximum product spacing, ordinary least squares, weighted least squares, Cramér-von Mises, percentile and Anderson–Darling estimation methods. Recently papers have been discussed the estimation methods for parameter of distribution modeling as [13, 14, 15].

4.1. Maximum Likelihood Estimation Method

Let the random samples $x_1, x_2, ..., x_n$ of size n be from the CBHE distribution. The maximum likelihood estimates (MLE) of the parameters of the distribution are obtained by first obtaining the log-likelihood function given as

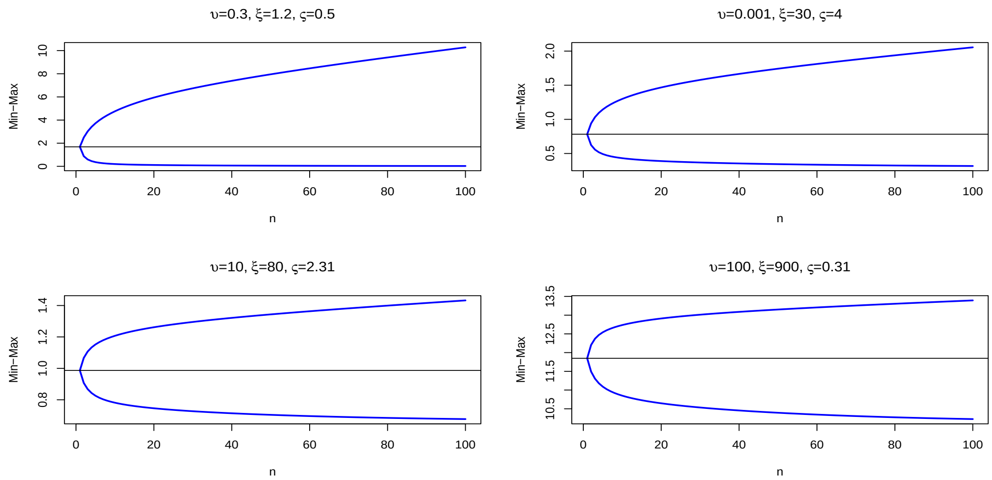

Figure 4. Minimum-Maximum Plots of Order Statistics of the CBHE Distribution

$$\ell = n\log(P\upsilon\varsigma\xi) + \sum_{i=1}^{n}\log\left(\frac{2+\varsigma x_i}{(1+\varsigma x_i)^2}\right) + (\xi-1)\sum_{i=1}^{n}\log\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)$$
$$-\varsigma\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi} + \upsilon\sum_{i=1}^{n}\left[1-e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}}\right].$$

Partially differentiating the likelihood function with respect to the parameters υ, ξ and ς, gives the score functions as

$$\frac{\partial\ell}{\partial\xi} = \frac{n}{\xi} + \sum_{i=1}^{n}\log\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right) + \sum_{i=1}^{n}\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}\log\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)$$
$$-\upsilon\sum_{i=1}^{n}\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi} e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}}\log\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right), \tag{4.1}$$

$$\frac{\partial\ell}{\partial\varsigma} = \frac{n}{\varsigma} - \sum_{i=1}^{n}x_i - \varsigma\sum_{i=1}^{n}\left(\frac{3+\varsigma x_i}{(1+\varsigma x_i)(2+\varsigma x_i)}\right) + (\xi-1)\sum_{i=1}^{n}\left(\frac{\varsigma e^{-\varsigma x_i}(2+\varsigma x_i)}{(1+\varsigma x_i)^2}\right)\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{-1}$$
$$+\xi\sum_{i=1}^{n}\left(\frac{\varsigma e^{-\varsigma x_i}(2+\varsigma x_i)}{(1+\varsigma x_i)^2}\right)\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi-1} + \upsilon\xi\sum_{i=1}^{n}\left(\frac{\varsigma e^{-\varsigma x_i}(2+\varsigma x_i)}{(1+\varsigma x_i)^2}\right)\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi-1} e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}}, \tag{4.2}$$

and

$$\frac{\partial\ell}{\partial\upsilon} = \frac{n}{\upsilon} + \frac{n(1-e)e^{\upsilon(1-e)}}{(1-e^{\upsilon(1-e)})} + \sum_{i=1}^{n}\left[1-e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}}\right]. \tag{4.3}$$

Setting the score functions in equations (4.1), (4.2) and (4.3) to zero and solving them simultaneously, the estimators of the parameters are obtained. Due to the non-linear nature of the score functions, numerical methods are employed to solve the equations.

4.2. Maximum Product Spacing Method

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be ordered random samples from CBHE distribution. Define the uniform spacing,

$$\Delta_i = F(x_{(i)}; \upsilon, \xi, \varsigma) - F(x_{(i-1)}; \upsilon, \xi, \varsigma),$$

where $F(x_{(0)}; \upsilon, \xi, \varsigma) = 0$, $F(x_{(n+1)}; \upsilon, \xi, \varsigma) = 1$ and $\sum_{i=0}^{n+1} \Delta_i = 1$. The maximum product spacing (MPS) estimators of the parameters are then obtained by maximizing the logarithm of the geometric mean of the spacing given by

$$LM = \frac{1}{n+1} \sum_{i=1}^{n+1} \log \Delta_i, i = 1, 2, ..., n+1$$

with respect to υ, ξ and ς.

4.3. Ordinary Least Squares Estimation Method

Given ordered random samples $x_{(1)}, x_{(2)}, ..., x_{(n)}$ of size n from the CBHE distribution with CDF $F(x)$. The ordinary least squares (OLS) estimators are obtained by minimizing the function

$$L = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right]^2 = \sum_{i=1}^{n} \left[\left(1 - P\left(1 - \exp\left\{ \upsilon \left(1 - e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\} \right) \right) - \frac{i}{n+1} \right]^2, \quad (4.4)$$

with respect to υ, ξ and ς. The parameter estimators are then obtained by simultaneously solving the resulting system of linear equations given by

$$\frac{\partial L}{\partial \xi} = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right] \Upsilon_1(x_i; \upsilon, \xi, \varsigma),$$

$$\frac{\partial L}{\partial \varsigma} = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right] \Upsilon_2(x_i; \upsilon, \xi, \varsigma)$$

and

$$\frac{\partial L}{\partial \upsilon} = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right] \Upsilon_3(x_i; \upsilon, \xi, \varsigma),$$

where

$$\Upsilon_1(x_i; \upsilon, \xi, \varsigma) = \left(\frac{2\upsilon}{1 - e^{\upsilon(1-e)}} \right) e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \exp\left\{ \upsilon \left(1 - e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\} \left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i} \right)^{\xi} \log\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i} \right), \quad (4.5)$$

$$\Upsilon_2(x_i; \upsilon, \xi, \varsigma) = \left(\frac{2\upsilon\xi}{1 - e^{\upsilon(1-e)}} \right) e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \exp\left\{ \upsilon \left(1 - e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\} \left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i} \right)^{\xi-1} \left(\frac{(2+\varsigma x_i)\, x_i e^{-\varsigma x_i}}{(1+\varsigma x_i)^2} \right), \quad (4.6)$$

and

$$\Upsilon_3(x_i; \upsilon, \xi, \varsigma) = \left(\frac{2}{(1 - e^{\upsilon(1-e)})^2} \right) (1-e) \left(e^{\upsilon(1-e)} \right) \left(1 - \exp\left\{ \upsilon \left(1 - e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\} \right) - \left(\frac{2}{1 - e^{\upsilon(1-e)}} \right) \left(1 - e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \exp\left\{ \upsilon \left(1 - e^{\left(1 - \frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\}. \quad (4.7)$$

4.4. Weighted Least Square Estimation Method

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ denote ordered random samples of size n with CDF, $F(x)$, from the CBHE distribution. The weighted least squares (WLS) estimators are then obtained by minimizing the function

$$W = \sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[F(x_{(i)}) - \frac{i}{n+1} \right],$$
$$= \sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[\frac{1}{1-e^{\upsilon(1-e)}} \left(1 - \exp\left\{ \upsilon \left(1 - e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\} \right) - \frac{i}{n+1} \right]^2, \tag{4.8}$$

with respect to υ, ξ and ς. Thus, solving the following system of equations will give the parameter estimators of the CBHE distribution:

$$\frac{\partial W}{\partial \xi} = \sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right] \Upsilon_1(x_i; \upsilon, \xi, \varsigma), \tag{4.9}$$

$$\frac{\partial W}{\partial \varsigma} = \sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right] \Upsilon_2(x_i; \upsilon, \xi, \varsigma) \tag{4.10}$$

and

$$\frac{\partial W}{\partial \upsilon} = \sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{i}{n+1} \right] \Upsilon_3(x_i; \upsilon, \xi, \varsigma). \tag{4.11}$$

where $\Upsilon_1(x_i; \upsilon, \xi, \varsigma)$, $\Upsilon_2(x_i; \upsilon, \xi, \varsigma)$ and $\Upsilon_3(x_i; \upsilon, \xi, \varsigma)$ are given by equations (4.5), (4.6), and (4.7) respectively.

4.5. Cramér-von Mises Estimation Method

Given the ordered random samples $x_{(1)}, x_{(2)}, ..., x_{(n)}$ from the CBHE distribution, the Cramér-von Mises (CVM) estimators are obtained by minimizing the function

$$C = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(x_{(i)}) - \frac{2i-1}{2n} \right]^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{1}{1-e^{\upsilon(1-e)}} \left(1 - \exp\left\{ \upsilon \left(1 - e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi}} \right) \right\} \right) - \frac{2i-1}{2n} \right]^2, \tag{4.12}$$

with respect to υ, ξ and ς. The CVM estimates are obtained by solving the system of linear equations obtained as follows

$$\frac{\partial C}{\partial \xi} = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{2i-1}{2n} \right] \Upsilon_1(x_i; \upsilon, \xi, \varsigma), \tag{4.13}$$

$$\frac{\partial C}{\partial \varsigma} = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{2i-1}{2n} \right] \Upsilon_2(x_i; \upsilon, \xi, \varsigma) \tag{4.14}$$

and

$$\frac{\partial C}{\partial \upsilon} = \sum_{i=1}^{n} \left[F(x_i; \upsilon, \xi, \varsigma) - \frac{2i-1}{2n} \right] \Upsilon_3(x_i; \upsilon, \xi, \varsigma), \tag{4.15}$$

where $\Upsilon_1(x_i; \upsilon, \xi, \varsigma)$, $\Upsilon_2(x_i; \upsilon, \xi, \varsigma)$ and $\Upsilon_3(x_i; \upsilon, \xi, \varsigma)$ are given by equations (4.5), (4.6), and (4.7) respectively.

4.6. Percentile Estimation Method

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ denote ordered random samples from the CBHE distribution with quantile function $Q(u_i)$, where $u_i = \frac{i}{n+1}$ is an unbiased estimator of $F(x_{(i)})$. The percentile (PC) estimators of the CBHE distribution are obtained by minimizing the function

$$P = \sum_{i=1}^{n} (x_{(i)} - Q(u_i))^2 = \sum_{i=1}^{n} \left(x_{(i)} - \left\{ \frac{1}{\varsigma} \left[W \left(e \left[1 - \left[\log \left[1 - \frac{1}{\upsilon} \log \left(1 - \frac{u_i}{P} \right) \right] \right]^{\frac{1}{\xi}} \right]^{-1} \right) - 1 \right] \right\} \right)^2, \quad (4.16)$$

with respect to υ, ξ and ς.

4.7. Anderson-Darling Estimation Method

The Anderson–Darling (AD) estimators of the parameters are obtained by minimizing the function

$$\mathrm{AD} = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[\log F(x_i; \upsilon, \xi, \varsigma) + \log \left(1 - F(x_{n+1-i}; \upsilon, \xi, \varsigma)\right) \right], \quad (4.17)$$

with respect to υ, ς and ξ. The AD estimators of the parameters for the CBHE distribution are then obtained by solving the nonlinear equations

$$\frac{\partial \mathrm{AD}}{\partial \xi} = -\frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[\frac{\Upsilon_1(x_i; \upsilon, \xi, \varsigma)}{F(x_i; \upsilon, \xi, \varsigma)} + \frac{\Upsilon_1(x_{n+1-i}; \upsilon, \xi, \varsigma)}{(1 - F(x_{n+1-i}; \upsilon, \xi, \varsigma))} \right] = 0, \quad (4.18)$$

$$\frac{\partial \mathrm{AD}}{\partial \upsilon} = -\frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[\frac{\Upsilon_2(x_i; \upsilon, \xi, \varsigma)}{F(x_i; \upsilon, \xi, \varsigma)} + \frac{\Upsilon_2(x_{n+1-i}; \upsilon, \xi, \varsigma)}{(1 - F(x_{n+1-i}; \upsilon, \xi, \varsigma))} \right] = 0, \quad (4.19)$$

and

$$\frac{\partial \mathrm{AD}}{\partial \varsigma} = -\frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[\frac{\Upsilon_1(x_3; \upsilon, \xi, \varsigma)}{F(x_i; \upsilon, \xi, \varsigma)} + \frac{\Upsilon_3(x_{n+1-i}; \upsilon, \xi, \varsigma)}{(1 - F(x_{n+1-i}; \upsilon, \xi, \varsigma))} \right] = 0, \quad (4.20)$$

where $\Upsilon_1(*; \upsilon, \xi, \varsigma)$, $\Upsilon_2(*; \upsilon, \xi, \varsigma)$ and $\Upsilon_3(*; \upsilon, \xi, \varsigma)$ are given by equations (4.5), (4.6), and (4.7), respectively.

4.8. Monte Carlo Simulations

In this section, Monte Carlo simulations are carried out to compare the performance of the proposed estimation techniques for estimating the parameters of the CBHE distribution. The estimation of the parameters is carried out by using the following procedure:

i. Generate random samples of size $n = 25, 50, 100, 250$ and 500 from the CBHE distribution using its quantile function in equation (3.1).
ii. Estimate the parameters of the distribution using MLE, MPS, OLS, WLS, AD, CVM and PC estimation methods.
iii. Steps i-ii are repeated for $N = 1000$ times.

iv. The average bias (AB) and root mean square error (RMSE) are computed for each parameter, υ, ξ and ς, using the following equations, respectively;

$$AB = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\Theta}_i - \Theta\right) \qquad \text{and} \qquad RMSE = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\Theta}_i - \Theta\right)^2}. \tag{4.21}$$

v. Steps i-iv are repeated for the parameter sets: $(\upsilon, \xi, \varsigma) = (0.8, 2.4, 1.2)$ and $(\upsilon, \xi, \varsigma) = (1.2, 0.9, 1.9)$.

The results of the simulation studies are presented in Figure 5 and Figure 6. Generally, it can be observed that the plots decrease with increasing sample size. This indicates that all the estimators are consistent. However, it can be observed that MLE estimates are generally lesser that the estimates of the other estimation methods. This suggests that MLE better estimates the parameters of CBHE distribution. Thus, MLE is used to estimate the parameters of the CBHE distribution for application purposes.

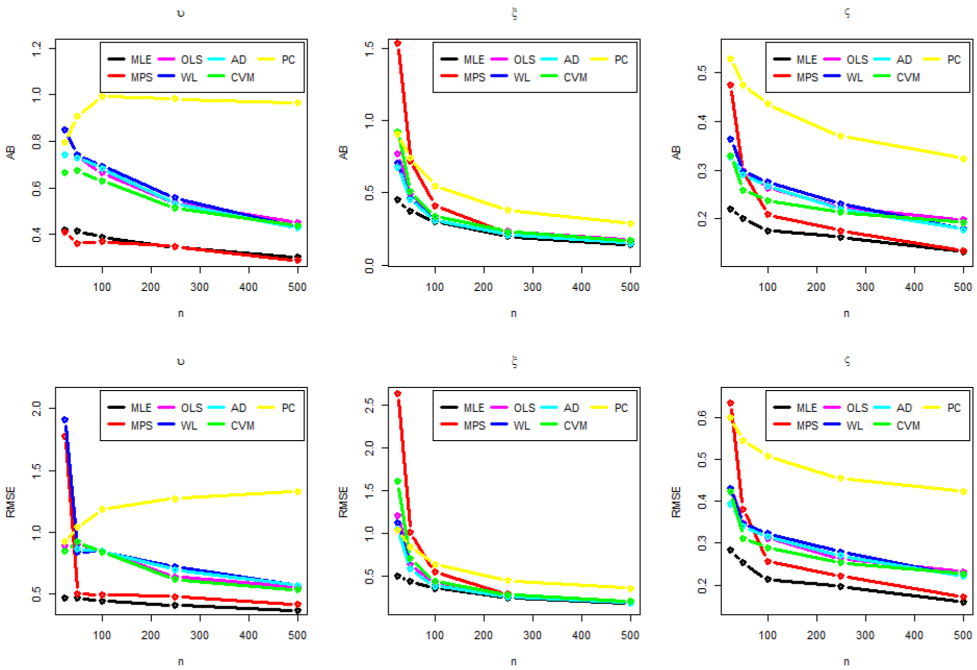

Figure 5. Plots of Simulated Results for $(\upsilon, \xi, \varsigma) = (0.8, 2.4, 1.2)$

5. Empirical Applications

The usefulness of CBHE distribution is demonstrated in this section using two real datasets. The performance of the CBHE distribution is compared with the performance of Chen exponential (CE) (Anzagra et al. [7]), exponential (E), BHE, Weibull (W) and WBXII distributions. The performance of the distributions are compared using Akaike information criterion (AIC), Bayesian information criterion (BIC), Cramér–von Mises (CVM), Anderson-Darling (AD) and Kolmogorov-Smirnov (KS) measures, with the corresponding p-values of CVM, AD and KS measures. The distribution with the

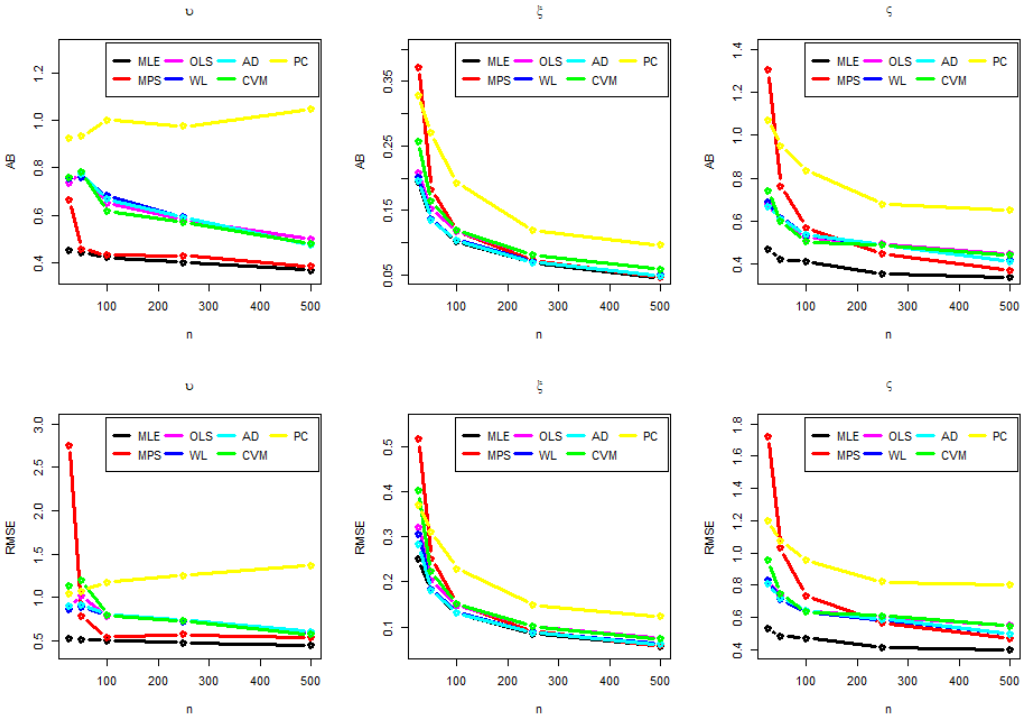

Figure 6. Plots of Simulated Results for $(\upsilon, \xi, \varsigma) = (1.2, 0.9, 1.9)$

least of these measures and the highest of the corresponding p-values of the goodness-of-fit measures is considered to be the best distribution for modeling a dataset.

5.1. Bladder Cancer Dataset

The first dataset consists of 128 monthly remission times of random samples of bladder cancer patients. The datasets can be found in Lee and Wang (2003) and are given as: 0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 1.46, 18.10, 11.79, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 13.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 12.07, 6.76, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

Figure 7 shows the kernel density and violin plots of the bladder cancer data. The shapes of the kernel density and violin plots indicate that the data is right skewed. This implies that the CBHE distribution can be used to model the dataset.

Table 2 shows the parameter estimates and the corresponding standard errors of the fitted distributions.

The information criteria and goodness-of-fit measures are presented in Table 3. It can be observed that CBHE distribution has the least of all the information criteria except for BIC measure, in which

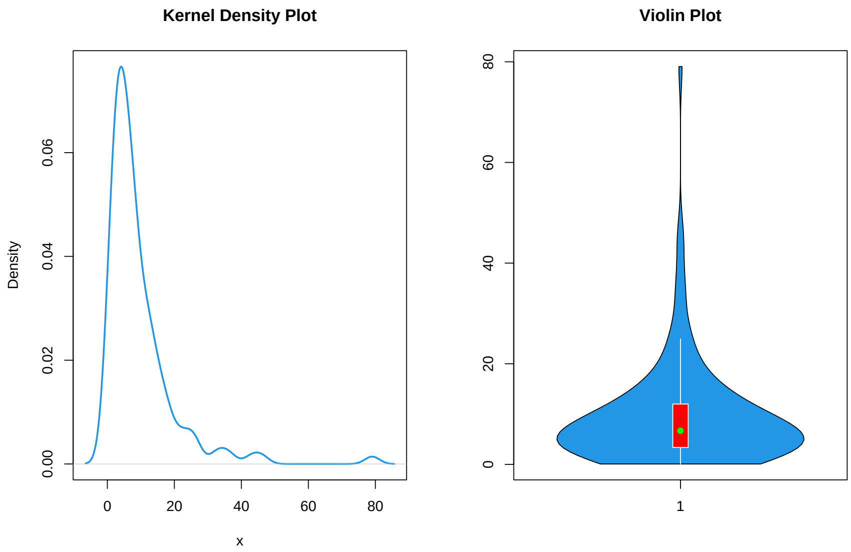

Figure 7. Kernel Density and Violin Plots of Bladder Cancer Dataset

Table 2. Parameter Estimates and Standard Errors of Fitted Distributions for Bladder Cancer Dataset

Distribution	$\hat{\upsilon}$		$\hat{\xi}$		$\hat{\varsigma}$		$\hat{\alpha}$	
	Estimate	**Std Error**	**Estimate**	**Std Error**	**Estimate**	**Std Error**	**Estimate**	**Std Error**
CBHE	2.1012	1.3427	1.427	0.1829	0.0428	0.0242		
CE	2.2836	1.0500	1.2979	0.1382	0.0660	0.0262		
E	0.1059	0.0094						
BHE	0.0618	0.0061						
W	1.0527	0.0680	0.0919	0.0188				
WBXII	0.8455	0.3128	0.2869	0.1493	1.7374	3.2358	2.1844	0.6472

the CBHE distribution has the second least measure. Also, the CBHE distribution has the least statistic and the highest corresponding p-values of the goodness-of-fit measures. This implies that the CBHE distribution is the best in modeling the bladder cancer dataset.

Table 3. Information Criteria and Goodness-of-fit Measures for Bladder Cancer Dataset

Distribution	AIC	BIC	CVM		AD		KS	
			Statistic	p-value	Statistic	p-value	Statistic	p-value
CBHE	828.1176	836.6737	0.0198	0.9973	0.1358	0.9994	0.0367	0.9953
CE	828.9441	837.5002	0.0313	0.9722	0.2070	0.9884	0.0413	0.9813
E	832.8104	835.6624	0.1697	0.3351	1.1273	0.2968	0.0832	0.3383
BHE	834.0375	836.8895	0.2901	0.1441	1.8172	0.1161	0.1060	0.1127
W	834.1968	839.9009	0.1380	0.4287	0.8744	0.4301	0.0663	0.6276
WBXII	832.5345	843.9426	0.0470	0.8950	0.3127	0.9282	0.0472	0.9381

Probability-probability (P-P) plots of the fitted distributions are obtained and given in Figure 8. It can be observed that the CBHE distribution provides a comparatively better fit to the dataset as its plots cluster more along the diagonal line.

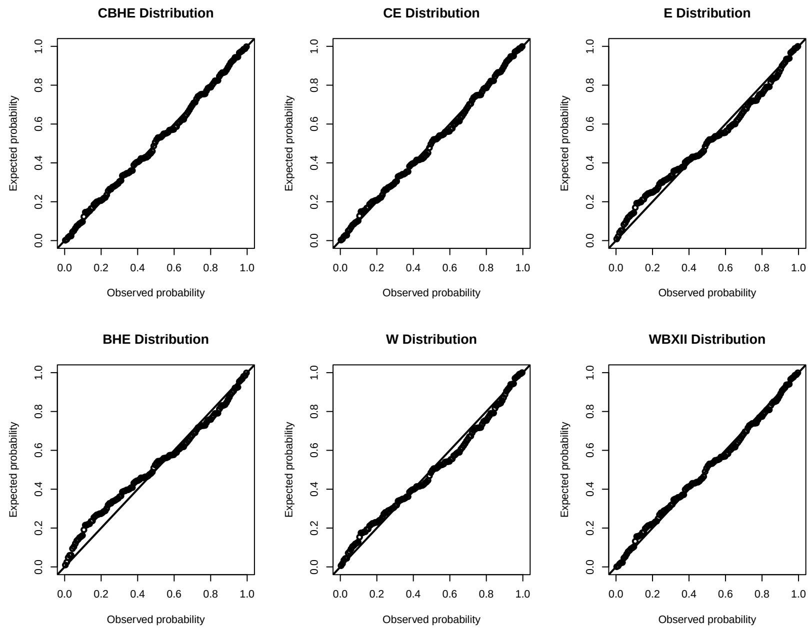

Figure 8. P-P Plots of Fitted Distributions for Bladder Cancer Dataset

5.2. Head and Neck Cancer Data

The second dataset consist of survival times of a group of patients suffering from head and neck cancer, and treated using a combination of radiotherapy and chemotherapy (RT+CT). The dataset was obtained from Efron [16] are given as: 12.20, 23.56, 23.74, 25.87, 31.98, 37, 41.35, 47.38, 55.46, 58.36, 63.47, 68.46, 78.26, 74.47, 81.43, 84, 92, 94, 110, 112, 119, 127, 130, 133, 140, 146, 155, 159, 173, 179, 194, 195, 209, 249, 281, 319, 339, 432, 469, 519, 633, 725, 817, 1776.

The kernel density and violin plots of the head and neck cancer dataset are shown in Figure 9. The shapes of the plots indicate that the data is right skewed. This implies that the CBHE distribution can be used to model the dataset.

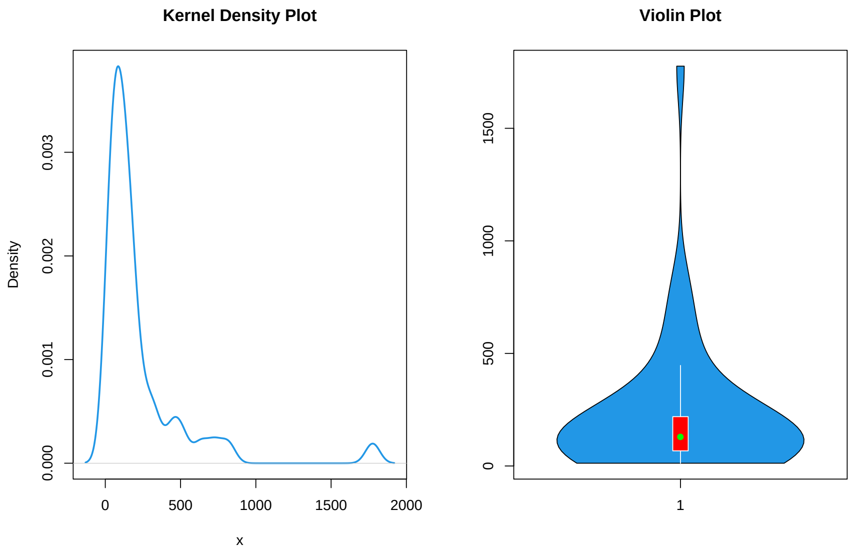

Figure 9. Kernel Density and Violin Plots of Head and Neck Cancer Dataset

Table 4 presents the estimated parameters of the fitted distributions with their corresponding standard errors.

Table 4. Parameter Estimates and Standard Errors of Fitted Distributions for Head and Neck Cancer Dataset

Distribution	$\hat{\upsilon}$		$\hat{\xi}$		$\hat{\varsigma}$		$\hat{\alpha}$	
	Estimate	**Std Error**	**Estimate**	**Std Error**	**Estimate**	**Std Error**	**Estimate**	**Std Error**
CBHE	1.7870	1.1368	1.4365	0.2984	0.0023	0.0013		
CE	1.8047	0.8345	1.2291	0.2293	0.0033	0.0012		
E	0.0045	0.0007						
BHE	0.0027	0.0005						
W	0.9234	0.1051	0.0071	0.0045				
WBXII	1.4435	0.5713	0.1044	0.0422	0.4876	0.5226	3.4630	0.4730

Table 5 presents the information criteria and goodness-of-fit measures of the fitted distributions. It can be observed that CBHE has the second least AIC measure after BHE distribution, and also has the third least BIC measure, after E and BHE distributions. This can be attributed to the fact CBHE distribution has more parameters than E and BHE distributions. However, the CBHE distribution has the least of CVM, AD and KS measures, with high p-values, which are significantly different from the other distributions. This implies that the CBHE distribution best models the dataset, comparatively.

Figure 10 shows P-P plots of the fitted distributions. It can be observed that the CBHE distribution has its plots clustering more around the diagonal as compared to the other distributions. Therefore, it can be concluded that the CBHE distribution best describes the neck and head cancer dataset as

Table 5. Information Criteria and Goodness-of-fit Measures for Head and Neck Cancer Dataset

Distribution	AIC	BIC	CVM		AD		KS	
			Statistic	p-value	Statistic	p-value	Statistic	p-value
CBHE	564.3278	569.6804	0.0363	0.9531	0.3544	0.8919	0.0673	0.9805
CE	566.0809	571.4334	0.0926	0.6248	0.5616	0.6841	0.1066	0.6599
E	566.0224	567.8065	0.1677	0.3409	0.9324	0.3942	0.1420	0.3073
BHE	563.9277	565.7119	0.1094	0.5427	0.7253	0.5372	0.1125	0.5941
W	567.7156	571.2840	0.1354	0.4388	0.8665	0.4348	0.1242	0.4684
WBXII	566.3276	573.4644	0.0840	0.6715	0.4765	0.7699	0.1121	0.5981

compared to the other competing distributions.

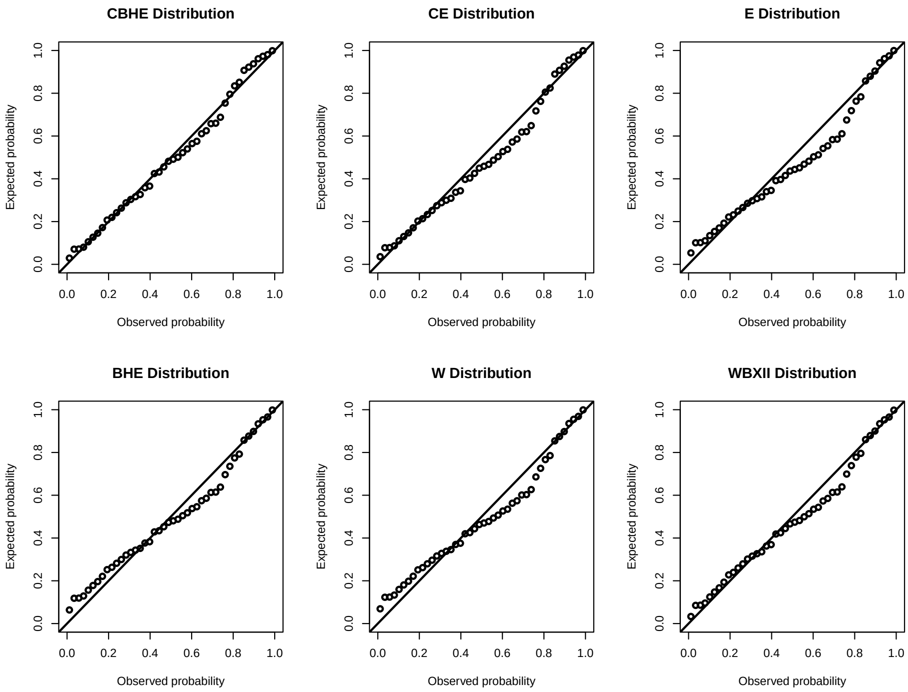

Figure 10. P-P Plots of Fitted Distributions for Head and Neck Cancer Dataset

6. Chen Burr-Hatke Exponential Regression Models

Regression models are useful in explaining the effects of some exogenous variables on a response variable. Therefore, in this section two regression models are proposed for response variables following the CBHE distribution.

6.1. CBHE Regression with Different Structures

Let $x_1, x_2, \ldots, x_n$ be random samples from the CBHE distribution. A CBHE regression model with different structures is established by relating the parameters υ and ξ to the covariates using an appropriate link function. That is

$$h(\upsilon_i) = \mathbf{Z}_i^T \boldsymbol{\alpha}, \qquad i = 1, 2, \ldots, n \tag{6.1}$$

and

$$h(\xi_i) = \mathbf{Z}_i^T \boldsymbol{\beta}, \qquad i = 1, 2, \ldots, n, \tag{6.2}$$

where $\mathbf{Z}_i^T = (z_{i1}, z_{i2}, \ldots, z_{ik})$ is an i^{th} vector of covariates, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^T$ are the vectors of regression coefficients, and $h(\bullet)$ is an appropriate link function. In this study the log, inverse, square root and inverse square root link functions are used.

The parameters ς, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, can be estimated by maximizing the log-likelihood function given as

$$\begin{aligned}\ell = n\log(P\varsigma) + \log\sum_{i=1}^{l}\log(\upsilon_i\xi_i) + \sum_{i=1}^{n}\log\left(\frac{2+\varsigma x_i}{(1+\varsigma x_i)^2}\right) + \sum_{i=1}^{n}\log\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi_i-1} \\ -\varsigma\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi_i} + \upsilon\sum_{i=1}^{n}\left[1-e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{\xi_i}}\right].\end{aligned} \tag{6.3}$$

6.2. Quantile Regression Model

Quantile regression models are appropriate in modeling data which are skewed or with outliers. Regression models which relate the covariates to the mean of the response variable fail to capture the variations in the response variable appropriately, as the mean fails to be the best measure of central tendency in the presence of outliers or when the data is skewed. This subsection presents CBHE quantile regression model. To achieve this, let $\lambda = Q(u), \lambda > 0$ and make ξ the subject. This gives

$$\xi = \left(\log\left\{\log\left[1-\frac{1}{\upsilon}\log\left(1-\frac{u}{P}\right)\right]\right\}\right)\left(\log\left\{\left(1-\frac{e^{-\varsigma\lambda}}{1+\varsigma\lambda}\right)\right\}\right)^{-1}. \tag{6.4}$$

Substituting equation (6.4) into the PDF of CBHE distribution, gives a re-parameterized PDF as

$$f(x) = P\upsilon\varsigma\frac{(2+\varsigma x)\log A}{(1+\varsigma x)^2\log R}H^{\frac{\log A}{\log R}-1}e^{-\varsigma x}\exp\left\{H^{\frac{\log A}{\log R}}\right\}\exp\left\{\upsilon\left[1-\exp\left\{H^{\frac{\log A}{\log R}}\right\}\right]\right\}, \tag{6.5}$$

where $A = \log\left[1-\frac{1}{\upsilon}\log\left(1-\frac{u}{P}\right)\right]$, $H = \left(1-\frac{e^{-\varsigma x}}{1+\varsigma x}\right)$ and $R = \left(1-\frac{e^{-\varsigma\lambda}}{1+\varsigma\lambda}\right)$.

Let $x_1, x_2, \ldots, x_n$ be random samples from the CBHE distribution. Then the quantile regression structure used in this study is given as

$$h(\lambda_i) = \mathbf{Z}_i^T\boldsymbol{\alpha}, \ i = 1, 2, \ldots, n, \tag{6.6}$$

where $\mathbf{Z}_i^T = (z_{i1}, z_{i2}, \ldots, z_{ik})$ is the ith vector of covariates, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)^T$ is a vector of regression coefficients and $h(\bullet)$ is an appropriate link function. In this study, the log link function is used for this regression model. Therefore,

$$\log(\lambda_i) = \mathbf{Z}_i^T\boldsymbol{\alpha}, \ i = 1, 2, \ldots, n \tag{6.7}$$

The parameters of the distribution can be obtained via maximum likelihood method by maximizing the log-likelihood function given as

$$\ell = n\log(P\upsilon\varsigma) + \sum_{i=1}^{n}\log D_i + \sum_{i=1}^{n}\log\left(\frac{2+\varsigma x_i}{(1+\varsigma x_i)^2}\right) + \sum_{i=1}^{n}\log\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{D_i-1} - \varsigma\sum_{i=1}^{n}x_i$$

$$+\sum_{i=1}^{n}\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{D_i}+\upsilon\sum_{i=1}^{n}\left[1-e^{\left(1-\frac{e^{-\varsigma x_i}}{1+\varsigma x_i}\right)^{D_i}}\right], \tag{6.8}$$

where $D_i = \left(\log\left\{\log\left[1-\frac{1}{\upsilon}\log\left(1-\frac{u}{P}\right)\right]\right\}\right)\left(\log\left\{\left(1-\frac{e^{-\varsigma e^{\mathbf{z}_i^T\alpha}}}{1+\varsigma e^{\mathbf{z}_i^T\alpha}}\right)\right\}\right)^{-1}$.

6.3. Residual Analysis

Residual analysis is essential to ascertain the adequacy of a fitted model. The Cox-Snell (Cox and Snell, [17]) residuals analysis is used in this study. Cox-Snell residuals are defined as

$$r_i = -\log\left(1-F\left(x_i;\hat{\boldsymbol{\theta}}\right)\right), i = 1, 2, \ldots, n, \tag{6.9}$$

where $\hat{\boldsymbol{\theta}}$ is a vector of estimated parameters. The residuals are expected to follow the standard exponential distribution if the model is adequate.

6.4. Application

The CBHE regression models developed are demonstrated in this subsection using real dataset. The effect of gender on the survival times (in years) until the onset of hypertension of 119 random samples and gender from the Bolgatanga Regional Hospital in the Upper East Region of Ghana is modeled. The dataset is obtained from Zamanah et al. [9]. The survival times, with gender (male = 1, female = 0) in brackets, is given as: 71(1), 5(1), 39(1), 62(1), 52(0), 71(0), 38(0), 56(1), 35(1), 69(1), 34(1), 71(1), 66(0), 70(1), 52(0), 37(0), 35(0), 71(1), 73(1), 19(0), 74(0), 74(1), 75(1), 51(0), 76(1), 49(0), 19(1), 76(0), 78(1), 76(0), 76(0), 49(1), 47(1), 48(0), 48(0), 46(0), 46(1), 46(1), 41(0), 40(0), 43(1), 45(0), 47(0), 47(0), 44(0), 45(1), 46(1), 42(1), 43(0), 42(0), 20(1), 28(0), 26(0), 60(0), 27(1), 24(0), 29(0), 60(1), 25(1), 60(1), 69(1), 36(1), 69(0), 69(1), 68(0), 68(0), 67(1), 67(0), 67(0), 52(0), 35(0), 66(0), 55(0), 66(1), 61(1), 61(0), 64(0), 64(0), 65(0), 65(0), 63(1), 63(1), 62(0), 39(1), 62(0), 62(0), 62(0), 59(1), 59(0), 59(1), 58(0), 58(0), 58(0), 18(1), 57(0), 57(0), 56(0), 56(0), 37(1), 53(0), 53(0), 53(0), 53(1), 54(1), 54(1), 66(0), 17(0), 50(0), 75(0), 51(0), 38(0), 52(1), 66(0), 4(1), 52(0), 55(0), 19(1), 58(1) ,73(0).

The following regression models are used:

A. $h(\upsilon_i) = \alpha_0 + \alpha_1 z_{i1}$ and $h(\xi_i) = \beta_0 + \beta_1 z_{i1}$, $i = 1, 2, \ldots, 119$ with the following link functions: log: $h(\bullet) = \log(\bullet)$; inverse: $h(\bullet) = (\bullet)^{-1}$; square root: $h(\bullet) = (\bullet)^{\frac{1}{2}}$ and inverse square: $h(\bullet) = (\bullet)^{-2}$ are used.

B. Quantile regression: $\log(\lambda_i) = \alpha_0 + \alpha_1 z_{i1}, i = 1, 2, \ldots, 119$ with the quantiles $u = 0.2, 0.25, 0.5$ and 0.75 are also used.

The models are compared using negative log-likelihood ($-\ell$), AIC and BIC measures.

Table 6 shows the parameter estimates, standard errors, p-values and information criteria of the fitted regression models. It can be observed that the covariate, that is gender, is significant for CBHE regression with different structures but not significant for the quantile regression model for all the quantiles at 5% significance level. It can also be observed that the parametric model with log link function has the least $-\ell$, AIC and BIC measures. This is followed by the quantile regression model

$u = 0.75$. For the best model, the covariate has a negative impact on the survival time until the onset of hypertension.

Figure 11 shows the P-P plots of the Cox-Snell residuals. It can be observed that the CBHE regression model with log link function has more of its plots clustering along the diagonal than the other regression models. This implies that it best models the dataset.

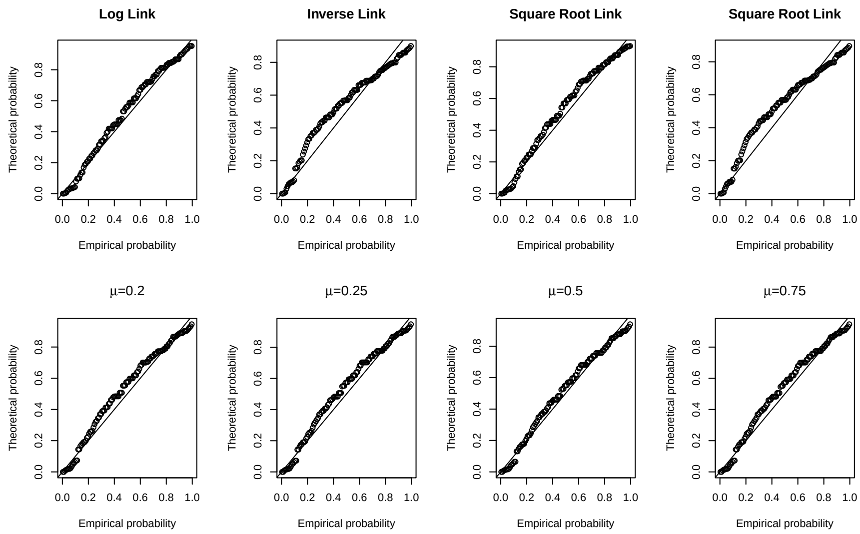

Figure 11. Cox-Snell P-P Plots of Fitted Regression Models

Table 6. Parameter Estimates and Information Criteria of Fitted Regression Models

Model			υ	ς	α_0	α_1	β_0	β_1	$-\ell$	AIC	BIC
A	log	Estimate		0.0011	9.9367	-2.9415	1.5531	-0.3698			
		Std Error		0.0002	0.0043	0.0082	0.0920	0.0235	500.7577	1011.515	1025.411
		p-value		< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001			
	Inverse	Estimate		0.0417	27.5417	33.3180	0.0671	0.0650			
		Std Error		0.0035	1.03×10^{-6}	2.99×10^{-7}	0.0165	0.0221	527.3497	1064.699	1078.595
		p-value		< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0033			
	Square root	Estimate		0.0057	9.0020	-3.4215	2.3877	-0.3274			
		Std Error		0.0009	0.2654	0.6591	0.1610	0.1163	506.2688	1022.538	1036.433
		p-value		< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0049			
	Inverse square	Estimate		0.0398	21.2696	21.8005	0.0047	0.0140			
		Std Error		0.0035	2.98×10^{-7}	1.03×10^{-7}	0.0023	0.0066	528.0003	1066.001	1079.896
		p-value		< 0.0001	< 0.0001	< 0.0001	0.0412	0.0345			
B	0.20	Estimate	68.8711	0.0046	3.6511	-0.0676					
		Std Error	1.61×10^{-5}	0.0007	0.0527	0.0681			508.2086	1024.417	1035.534
		p-value	< 0.0001	< 0.0001	< 0.0001	0.3206					
	0.25	Estimate	99.0519	0.0040	3.7213	-0.0600					
		Std Error	8.63×10^{-6}	0.0006	0.0486	0.0647			507.5277	1023.055	1034.172
		p-value	< 0.0001	< 0.0001	< 0.0001	0.3542					
	0.50	Estimate	98.8646	0.0041	3.9842	-0.0523					
		Std Error	7.01×10^{-6}	0.0006	0.0382	0.0575			507.759	1023.519	1034.635
		p-value	< 0.0001	< 0.0001	< 0.0001	0.3634					
	0.75	Estimate	107.2043	0.0039	4.1682	-0.0474					
		Std Error	6.22×10^{-6}	0.0006	0.0360	0.0528			507.3871	1022.774	1033.891
		p-value	< 0.0001	< 0.0001	< 0.0001	0.3686					

7. Conclusion

In this study, a new lifetime distribution, called Chen Burr-Hatke exponential (CBHE) distribution, was developed. Various plots of the density function showed that the distribution can exhibit increasing, decreasing, right-skewed and left-skewed shapes, whiles plots of the hazard rate function showed that the distribution can exhibit increasing, decreasing, and upside-down bathtub shapes. Several statistical properties including quantile function, moments, order statistics and inequality measures were derived. The parameters of the distribution were estimated using seven estimation methods. The estimators were all consistent, however, maximum likelihood was observed to better estimate the parameters of the distribution via simulation studies. Associated regression models of the CBHE distribution were developed. The applications and usefulness of the CBHE distribution and its regression models were demonstrated for real datasets. The results showed that the distribution performed better than the competing distributions. Therefore, it can be concluded that the CBHE distribution and its regression models can serve as alternative models to describing datasets from biomedical sciences and other fields, which exhibit various characteristics. The CBHE distribution is therefore recommended for use in modeling such datasets.

References

1. Da Silva, R. V., Gomes-Silva, F., Ramos, M. W. A., & Cordeiro, G. M. (2015). The exponentiated Burr XII Poisson distribution with application to lifetime data. *International Journal of Statistics and Probability*, **4**(4): 112-131.
2. Afify, A. Z., Cordeiro, G. M., Ortega, E. M., Yousof, H. M., and Butt, N. S. (2018). The four-parameter Burr XII distribution: properties, regression model, and applications. *Communications in Statistics-Theory and Methods*, **47**(11): 2605–2624.
3. Alizadeh, M., Lak, F., Rasekhi, M., Ramires, T. G., Yousof, H. M., & Altun, E. (2018). The odd log-logistic Topp–Leone G family of distributions: heteroscedastic regression models and applications. *Computational Statistics*, **33**(3): 1217-1244.
4. Abouelmagd, T. H. M. (2018). The Logarithmic Burr-Hatke Exponential Distribution for Modeling Reliability and Medical Data. *International Journal of Statistics and Probability*, **7**(5): 73-85.
5. Yadav, A.S., Altun, E. & Yousof, H.M.(2021). Burr–Hatke Exponential Distribution: A Decreasing Failure Rate Model, Statistical Inference and Applications. *Annals of Data Science*, **8**(2): 241–260.
6. Nasir, M. A., Tahir, M. H., Chesneau, C., Jamal, F., & Shah, M. A. A. (2020). The odds generalized gamma-G family of distributions: Properties, regressions and applications. *Statistica*, **80**(1): 3-38.
7. Anzagra, L., Sarpong, S., & Nasiru, S. (2020). Chen-G class of distributions. *Cogent Mathematics & Statistics*, **7**(2):1721.
8. Afify, A. Z., Aljohani, H. M., Alghamdi, A. S., Gemeay, A. M., & Sarg, A. M. (2021). A new two-parameter burr-hatke distribution: properties and bayesian and non-bayesian inference with applications. *Journal of Mathematics*, 2021.
9. Zamanah, E., Nasiru, S., & Luguterah, A. (2022). Harmonic Mixture Weibull-G Family of Distributions: Properties, Regression and Applications to Medical Data. *Computational and Mathematical Methods*, 2022.

10. Kharazmi, O., Nik, A. S., Hamedani, G. G., & Altun, E. (2022). Harmonic Mixture-G Family of Distributions: Survival Regression, Simulation by Likelihood, Bootstrap and Bayesian Discussion with MCMC Algorithm. *Austrian Journal of Statistics*, **51**(2): 1-27.

11. Rasekhi, M., Altun, E., Alizadeh, M., & Yousof, H. M. (2022). The Odd Log-Logistic Weibull-G Family of Distributions with Regression and Financial Risk Models. *Journal of the Operations Research Society of China*, **10**(1): 133-158.

12. Gastwirth J (1972). The estimation of the Lorenz curve and Gini index. *Rev Econ Stat,* **54**(3): 306–316.

13. Hamdy, A., & Almetwally, E. M. (2023). Bayesian and Non-Bayesian Inference for The Generalized Power Akshaya Distribution with Application in Medical. Computational Journal of Mathematical and Statistical Sciences, **2**(1), 31-51.

14. Ocloo, S. K., Brew, L., Nasiru, S., & Odoi, B. (2023). On the Extension of the Burr XII Distribution: Applications and Regression. Computational Journal of Mathematical and Statistical Sciences, **2**(1), 1-30.

15. Tolba, A. (2022). Bayesian and Non-Bayesian Estimation Methods for Simulating the Parameter of the Akshaya Distribution. Computational Journal of Mathematical and Statistical Sciences, **1**(1), 13-25.

16. Efron, B. (1988). Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association*. **83**(402): 414-425.

17. Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the RoyalStatistical Society, Series B*, **30**: 248-275.