Mikailalsys Journal of
Mathematics and Statistics

e-ISSN : 3030-816X
p-ISSN : 3030-8399

Index : Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/MJMS.v2i3.3840

Novel Extended Weibull Regression Model for Investigating the Survival Times of Breast Cancer Patients

Ahmed Abdulkadir[1], Obinna Damian Adubisi[2], R. M. Madaki[3]
[1,3]Abubakar Tafawa Balewa University, Nigeria; [2]Federal University Wukari, Nigeria
aabdulkadir@atbu.edu.ng; adubisiobinna@fuwukari.edu.ng

Article Info:

Submitted:	Revised:	Accepted:	Published:
Sep 1, 2024	Sep 11, 2024	Sep 14, 2024	Sep 17, 2024

Abstract

The new five-parameter alpha power generalized odd generalized exponentiated Weibull distribution is introduced, and some of its structural properties are derived. Its parameters are estimated by maximum likelihood, and a simulation study examines the accuracy of the estimates. A regression model is constructed based on the logarithm of the proposed distribution to investigate the survival times of breast cancer patients in Bauchi State, Nigeria. The applicability and flexibility of the novel model is proven by means of cancer dataset.

Keywords: Alpha-power transformation; Breast cancer; Censored data; Maximum likelihood; Regression model

INTRODUCTION

Cancer is an ailment triggered by the unrestrained separation of abnormal cells in a portion of the human body most times. It starts when cells in a portion of the body start to grow wildly due to genetic vicissitudes that weaken their usual evolution. The cancer mostly diagnosed among all women is breast cancer. Breast cancer represent one-quarter of all cancerous issues detected in women worldwide and the foremost cause of cancer demises among women universally. Breast cancer is an assembly of ailments in which cells in breast tissue transform and split uncontrolled, typically resulting in a lump or mass. Most breast cancers begin in the lobules (milk glands) or in the ducts that connect the lobules to the nipple (ACS, 2019). Breast cancer cases are often characterized by late diagnostics, considered as the primary cause of cancer related deaths among women in Nigeria (Nnaka et al., 2022). Nabegu et al. (2023) assessed the length of life of breast cancer patients and the prognostic factors associated with the patients' survival using medical records and pathological variables of women with breast cancer from Aminu Kano Teaching hospital (AKTH), Nigeria. Nnaka et al. (2022) studied the five years retrospective review of age distribution and histo-epidemiological profile of breast cancer cases in Bauchi state, Nigeria. However, these studies only informed on the prevalence rate of breast cancer based on associated factors used.

Furthermore, the quality of any statistical analysis carried out to model a dataset depends on the statistical distribution selected. Given that various datasets are characterized by different features, the familiar classical distributions are not always adequate to describe the definite behaviour of these datasets. Hence, several transformed, augmented, composite, and mixed distributions have been developed and applied in these datasets from various fields. However, there are still many vital issues that cannot be explained by the existing distributions, so we need more flexible and consistent distribution for these issues. One of the most significant and current issues that has piqued our attention is breast cancer cases among women. Several researchers had carried out research on breast cancer, such as Adamu et al. (2019), Feleke et al. (2022), Misganaw et al. (2023). The limitation of these studies is that they were only carried out from the survival analysis point of view.

Hence, this research is set out to investigate the survival time of breast cancer patients using a novel log-alpha power generalized odd generalized exponentiated Weibull (log-APGOGEW) regression model. This can be achieved by developing a regression model with log-APGOGEW error distribution. The novel log-APGOGEW regression model will use socio-demographic variables and clinical factors such as age, clinical stage and body mass index (BMI) as covariates in investigating the survival times (time to death) of breast cancer patients.

METHODS

Model Genesis

The cumulative distribution function (CDF) of the Weibull distribution (Weibull, 1951) is defined as

$$\mathbf{G}(\mathbf{x};\eta,\varphi)=1-\mathbf{e}^{-\left(\mathbf{x}/\eta\right)^{\varphi}},\quad \eta,\varphi>0, \tag{1}$$

and the corresponding probability density function (PDF) is

$$\mathbf{g}(\mathbf{x};\eta,\varphi)=\frac{\varphi}{\eta}\left(\frac{\mathbf{x}}{\eta}\right)^{\varphi-1}\mathbf{e}^{-\left(\mathbf{x}/\eta\right)^{\varphi}} \tag{2}$$

The CDF and PDF of the APGOGE-G class developed by Abdulkadir et al. (2024) are specified as

$$\mathrm{F}(\mathrm{x};\alpha,\beta,\tau,\zeta)=\begin{cases}\dfrac{\alpha^{\left\{1-\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}\right\}^{\beta}}-1}{\alpha-1}, & \alpha>0,\ \alpha\neq1,\beta,\tau>0\\[2ex] \left\{1-\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}\right\}^{\beta}, & \alpha=1,\end{cases} \tag{3}$$

and

$$\mathrm{f}(\mathrm{x};\alpha,\beta,\tau,\zeta)=\begin{cases}\dfrac{\tau\beta\log(\alpha)\mathrm{g}(\mathrm{x};\zeta)\mathrm{G}(\mathrm{x};\zeta)^{\tau-1}\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}}{(\alpha-1)\left[1-\mathrm{G}(\mathrm{x};\zeta)^{\tau}\right]^{2}}\left\{1-\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}\right\}^{\beta-1}\alpha^{\left\{1-\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}\right\}^{\beta}}, & \alpha>0,\ \alpha\neq1,\\[2ex] \dfrac{\tau\beta\mathrm{g}(\mathrm{x};\zeta)\mathrm{G}(\mathrm{x};\zeta)^{\tau-1}}{\left[1-\mathrm{G}(\mathrm{x};\zeta)^{\tau}\right]^{2}}\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}\left\{1-\mathrm{e}^{\frac{-\mathrm{G}(\mathrm{x};\zeta)^{\tau}}{\bar{\mathrm{G}}(\mathrm{x};\zeta)^{\tau}}}\right\}^{\beta-1}, & \alpha=1,\end{cases} \tag{4}$$

where ζ is the parameter vector of $G(.)$

By inserting Equations (1) and (2) into Equations (3) and (4) for $\alpha>0,\alpha\neq1$, the CDF and PDF of the random variable $X \sim APGOGEW(\alpha,\beta,\tau,\eta,\varphi)$ is specified as

$$\mathrm{F}(\mathrm{x}) = \frac{\alpha^{\left\{1-\mathrm{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta}} - 1}{\alpha - 1}, \tag{5}$$

and

$$\mathrm{f}(\mathrm{x}) = \frac{\tau\beta\varphi\eta^{-1}\log(\alpha)\left(\frac{\mathrm{x}}{\eta}\right)^{\varphi-1} \mathrm{e}^{-\left(x/\eta\right)^{\varphi}} (\Phi)^{\tau-1} \mathrm{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}}{(\alpha-1)\left[1-(\Phi)^{\tau}\right]^{2}} \left\{1-\mathrm{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta-1} \alpha^{\left\{1-\mathrm{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta}} \tag{6}$$

where $\Phi = 1 - e^{-\left(x/\eta\right)^{\varphi}}$ and the hazard rate function (HRF) is specified as

$$\mathrm{f}(\mathrm{x}) = -\frac{\tau\beta\varphi\eta^{-1}\log(\alpha)\left(\frac{\mathrm{x}}{\eta}\right)^{\varphi-1} \mathrm{e}^{-\left(x/\eta\right)^{\varphi}} (\Phi)^{\kappa-1} \mathrm{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}}{\left[1-(\Phi)^{\kappa}\right]^{2}} \left\{1-\mathrm{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta-1} \tag{7}$$

The PDF plots of X in Figure 1 depict some interesting shapes such as symmetric, increasing, decreasing, right-skewed. Likewise, the HRF plots of X in Figure 2 depicts increasing, decreasing and bathtub shapes.

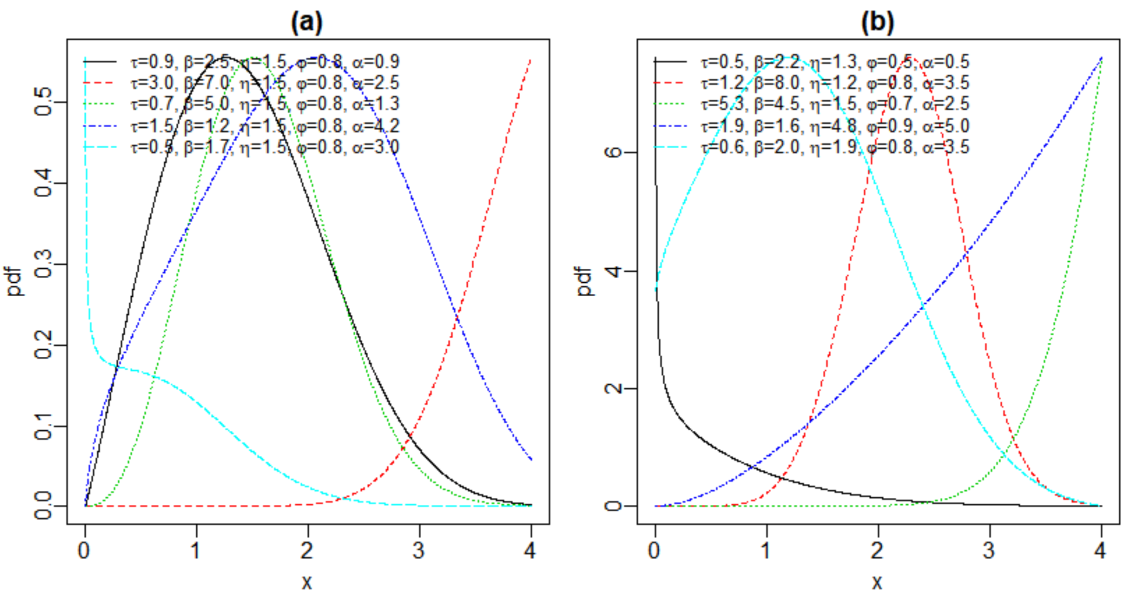

Figure 1: PDF plots of $X \sim APGOGEW(\alpha, \beta, \tau, \eta, \varphi)$.

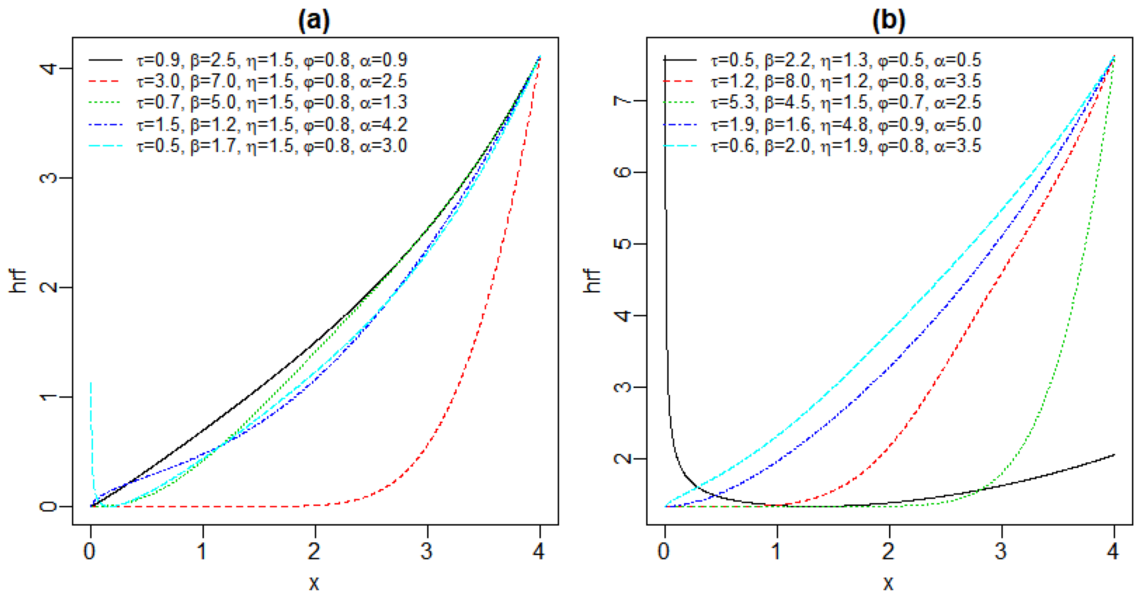

Figure 2: HRF plots of $X \sim APGOGEW(\alpha,\beta,\tau,\eta,\varphi)$.

Structural Properties

The quantile function of X derived by inverting Equation (5) is specified as

$$Q(u)=\eta\left(-\log\left\{1-\left[\frac{-\log\left(1-\left\{\frac{\log\left[1+u(\alpha-1)\right]}{\log\alpha}\right\}^{\frac{1}{\beta}}\right)}{1-\log\left(1-\left\{\frac{\log\left[1+u(\alpha-1)\right]}{\log\alpha}\right\}^{\frac{1}{\beta}}\right)}\right]^{\frac{1}{\tau}}\right\}\right)^{\frac{1}{\varphi}} \tag{8}$$

Hence, pseudo-random numbers can be generated from the APGOGEW distribution using Equation (8).

The r[th] raw-moment (RM) of X is specified as $\mu_r'=E(X^r)=\int_0^\infty x^r f(x)\,dx$, then utilizing Equation (6), the rth moment of the APGOGEW distribution can be specified as

$$\mu_r'=\int_0^\infty x^r\frac{\tau\beta\varphi\eta^{-1}\log(\alpha)\left(\frac{x}{\eta}\right)^{\varphi-1}e^{-\left(x/\eta\right)^{\varphi}}(\Phi)^{\tau-1}e^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}}{(\alpha-1)\left[1-(\Phi)^{\tau}\right]^2}\left\{1-e^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta-1}\alpha^{\left\{1-e^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta}}dx. \tag{9}$$

The integral form in Equation (9) can be solved numerically utilizing R-programme. However, an analytical form for the RM can be derived by applying the power series $\alpha^{z}=\sum_{i=0}^{\infty}\frac{(\log\alpha)^{i}}{i!}z^{i}$ and generalized Binomial series theorem on Equation (9), we have

$$\mu_r'=\frac{\tau\beta\varphi\eta^{-\varphi}}{(\alpha-1)}\sum_{i,j,k=0}^{\infty}\frac{(-1)^{i+k}(\log\alpha)^{\alpha+1}(j+1)^{k}}{i!k!}\binom{\beta(\alpha+1)-1}{j}\times\int_0^{\infty}x^{r+\varphi-1}e^{-\left(x/\eta\right)^{\varphi}}\frac{(\Phi)^{\tau(k+1)-1}}{\left[1-(\Phi)^{\tau}\right]^{k+2}}dx. \tag{10}$$

Applying the Binomial theorem on $(\Phi)^{\tau(k+1)-1}\Big/\left[1-(\Phi)^{\tau}\right]^{k+2}$ gives

$$\mu_r'=\frac{\tau\beta\varphi\eta^{-\varphi}}{(\alpha-1)}\sum_{i,j,k,g=0}^{\infty}\frac{(-1)^{i+k+g}(\log\alpha)^{\alpha+1}(j+1)^{k}}{i!k!}\binom{\beta(\alpha+1)-1}{j}\times\binom{-(k+2)}{g}\int_0^{\infty}x^{r+\varphi-1}e^{-\left(x/\eta\right)^{\varphi}}(\Phi)^{\tau(k+g+1)-1}dx. \tag{11}$$

Again, utilizing the Binomial theorem on $(\Phi)^{\tau(k+g+1)-1}$ gives

$$\mu_r'=\frac{\tau\beta\varphi\eta^{-\varphi}}{(\alpha-1)}\sum_{i,j,k,g,h=0}^{\infty}\frac{(-1)^{i+k+g+h}(\log\alpha)^{\alpha+1}(j+1)^{k}}{i!k!}\binom{\beta(\alpha+1)-1}{j}\times\binom{-(k+2)}{g}\binom{\tau(k+g+1)-1}{h}\int_0^{\infty}x^{r+\varphi-1}e^{-(h+1)\left(x/\eta\right)^{\varphi}}dx. \tag{12}$$

Therefore, the rth RM of the APGOGEW distribution is specified as

$$\mu_r'=\sum_{h=0}^{\infty}\Pi_h\frac{\Gamma\left(\frac{r}{\varphi}+1\right)}{\left[(h+1)\eta^{-\varphi}\right]^{\frac{r}{\varphi}+1}} \tag{13}$$

where $\Gamma(\cdot)$ is the gamma function and

$$\Pi_r=\frac{\tau\beta\eta^{-\varphi}}{(\alpha-1)}\sum_{i,j,k,g=0}^{\infty}\frac{(-1)^{i+k+g+h}(\log\alpha)^{\alpha+1}(j+1)^{k}}{i!k!}\binom{\beta(\alpha+1)-1}{j}\binom{-(k+2)}{g}\binom{\tau(k+g+1)-1}{h}$$

Table 1. Numerical moments of the APGOGEW.

μ'_r	**COM1**	**COM2**	**COM3**
μ'_1	0.05796	0.00912	0.00099
μ'_2	0.02942	0.00573	0.00070
μ'_3	0.01850	0.00405	0.00053
μ'_4	0.01307	0.00308	0.00042
μ'_r	**COM4**	**COM5**	**COM6**
μ'_1	0.25510	0.03422	0.00173
μ'_2	0.17820	0.02852	0.00155
μ'_3	0.13620	0.02442	0.00141
μ'_4	0.10990	0.02134	0.00128

Table 1 reports the four RM of X from Equation (9) for selected values of the APGOGEW parameters combinations (Com): **Com1:** $\alpha = 0.5, \beta = 0.5, \tau = 0.5, \eta = 1.5, \varphi = 1.5$, **Com2:** $\alpha = 0.5, \beta = 0.5, \tau = 0.5, \eta = 2.5, \varphi = 2.5$, **Com3:** $\alpha = 0.5, \beta = 0.5, \tau = 0.5, \eta = 3.5, \varphi = 3.5$, **Com4:** $\alpha = 1.5, \beta = 1.5, \tau = 1.5, \eta = 0.5, \varphi = 0.5$, **Com5:** $\alpha = 2.5, \beta = 2.5, \tau = 2.5, \eta = 0.5, \varphi = 0.5$, **Com6:** $\alpha = 3.5, \beta = 3.5, \tau = 3.5, \eta = 0.5, \varphi = 0.5$.

The variance, skewness and kurtosis values of X are without difficulty obtained from the four RM. The 3D plots depicted in Figure 3, shows that the skewness and kurtosis can be decreasing as the values of η, φ increases with fixed $\alpha = \beta = \tau = 0.5$.

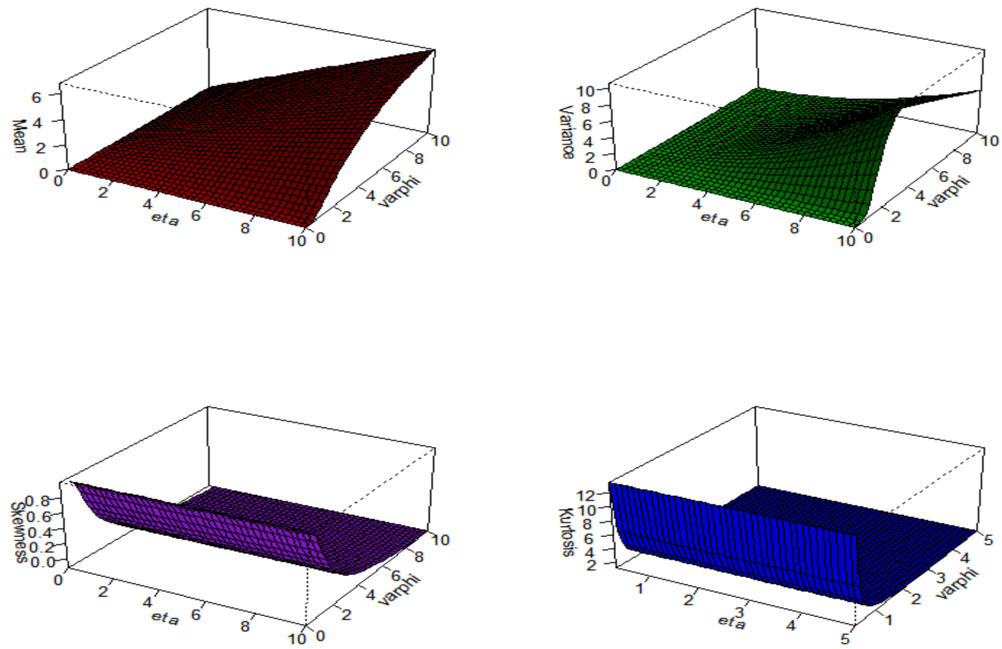

Figure 3: 3D plots of mean, variance, skewness and kurtosis for the APGOGEW

More so, the r^th^ incomplete moment (IM) of X, say $\vartheta_{\mathbf{r}}(\mathbf{t})=\int_0^{\mathbf{t}}\mathbf{x}^{\mathbf{r}}\mathbf{f}(\mathbf{x})\mathbf{dx},$ is specified as

$$\vartheta_{\mathbf{r}}(\mathbf{t})=\int_0^{\mathbf{t}}\mathbf{x}^{\mathbf{r}}\frac{\tau\beta\varphi\eta^{-1}\log(\alpha)\left(\frac{\mathbf{x}}{\eta}\right)^{\varphi-1}\mathbf{e}^{-\left(\mathbf{x}/\eta\right)^{\varphi}}(\Phi)^{\tau-1}\mathbf{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}}{(\alpha-1)\left[1-(\Phi)^{\tau}\right]^{2}}\left\{1-\mathbf{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta-1}\alpha^{\left\{1-\mathbf{e}^{\frac{-(\Phi)^{\tau}}{1-(\Phi)^{\tau}}}\right\}^{\beta}}\mathbf{dx}. \quad (14)$$

Utilizing the similar steps that lead to Equation (13), we have

$$\vartheta_{\mathbf{r}}(\mathbf{t})=\frac{\tau\beta\varphi\eta^{-\varphi}}{(\alpha-1)}\sum_{\mathbf{i,j,k,g,h}=0}^{\infty}\frac{(-1)^{\mathbf{i+k+g+h}}(\log\alpha)^{\alpha+1}(\mathbf{j}+1)^{\mathbf{k}}}{\mathbf{i}!\mathbf{k}!}\binom{\beta(\alpha+1)-1}{\mathbf{j}} \times\binom{-(\mathbf{k}+2)}{\mathbf{g}}\binom{\tau(\mathbf{k}+\mathbf{g}+1)-1}{\mathbf{h}}\int_0^{\mathbf{t}}\mathbf{x}^{\mathbf{r}+\varphi-1}\mathbf{e}^{-(\mathbf{h}+1)\left(\mathbf{x}/\eta\right)^{\varphi}}\mathbf{dx}. \quad (15)$$

Hence, the IM for the APGOGEW is specified as

$$\vartheta_{\mathbf{r}}(\mathbf{t})_{\mathbf{r}}=\sum_{\mathbf{h}=0}^{\infty}\Pi_{\mathbf{h}}\frac{\gamma\left(\frac{\mathbf{r}}{\varphi}+1,(\mathbf{h}+1)\eta^{-\varphi}\mathbf{t}^{\varphi}\right)}{\left[(\mathbf{h}+1)\eta^{-\varphi}\right]^{\frac{\mathbf{r}}{\varphi}+1}} \quad (16)$$

where

$$\Pi_r = \frac{\tau\beta\eta^{-\varphi}}{(\alpha-1)}\sum_{i,j,k,g=0}^{\infty}\frac{(-1)^{i+k+g+h}(\log\alpha)^{\alpha+1}(j+1)^k}{i!k!}\binom{\beta(\alpha+1)-1}{j}\binom{-(k+2)}{g}\binom{\tau(k+g+1)-1}{h}$$

where $\gamma(\cdot)$ is the lower incomplete gamma function.

The first IM $\vartheta_1(t)$ is utilized in the computation of important statistical measures such as the mean deviation about the mean and median of X. These are specified as $\ell_1 = 2\mu_1'F(\mu_1') - 2\vartheta_1(\mu_1')$ and $\ell_2 = \mu_1' - 2\vartheta_1(M)$, respectively. where μ_1 is the mean derived by setting $r = 1$ in Equation (13) and M is obtained by setting $u = 0.5$ in Equation (8).

Estimation

Let $X \sim APGOGEW(\xi)$, where $\xi = (\tau, \beta, \alpha, \eta, \varphi)^T$ is the vector of unknown parameters. The log-likelihood function $\ell(\xi)$ is specified as

$$\begin{aligned}\ell(\xi) &= n\log\tau + n\log\beta + n\log\varphi - n\log\eta + n\log(\log\alpha) - \log(\alpha-1) \\ &+(\varphi-1)\sum_{i=1}^{n}\log\left(\frac{x_i}{\eta}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\eta}\right)^{\varphi} + (\tau-1)\sum_{i=1}^{n}\log\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right) \\ &-2\sum_{i=1}^{n}\log\left(1-\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}\right) - \sum_{i=1}^{n}\frac{\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}}{1-\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}} \\ &+(\beta-1)\sum_{i=1}^{n}\log\left(1-e^{-\frac{\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}}{1-\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}}}\right) + \sum_{i=1}^{n}\left(1-e^{-\frac{\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}}{1-\left(1-e^{-\left(x/\eta\right)^{\varphi}}\right)^{\tau}}}\right)^{\beta}\log(\alpha).\end{aligned} \tag{17}$$

The maximum likelihood estimate (MLE) $\hat{\xi}$ of ξ is determined by maximizing Equation (17) numerically using R-programme (*optim function*).

Regression

Given that X has the PDF specified in Equation (6) and $Y = \log X$, then setting $\eta = e^{\mu}$ and $\varphi = \sigma^{-1}$. The log-APGOGEW (LAPGOGEW) density of Y $\left(\text{for } y \in \mathbb{R}\right)$ can be specified as

$$f\left(y;\tau,\beta,\alpha,\mu,\sigma\right)=\frac{\tau\beta\log\left(\alpha\right)e^{\frac{y-\mu}{\sigma}}e^{-e^{\frac{y-\mu}{\sigma}}}\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau-1}e^{\frac{-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}{1-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}}}{\sigma\left(\alpha-1\right)\left[1-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}\right]^{2}}$$

$$\times\left\{1-e^{\frac{-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}{1-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}}\right\}^{\beta-1}\alpha^{\left\{1-e^{\frac{-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}{1-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}}\right\}^{\beta}} \qquad (18)$$

where $\tau,\beta,\alpha,\sigma > 0$ and $\mu \in \mathbb{R}$. Thus, if $X \sim \text{APGOGEW}\left(\tau,\beta,\alpha,\eta,\varphi\right)$, then $Y = \log\left(X\right) \sim \text{LAPGOGEW}\left(\tau,\beta,\alpha,\mu,\sigma\right)$. The survival function corresponding to Equation (18) is specified as

$$S\left(z;\tau,\beta,\alpha,\mu,\sigma\right)=1-\frac{\alpha^{\left\{1-e^{\frac{-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}{1-\left(1-e^{-e^{\frac{y-\mu}{\sigma}}}\right)^{\tau}}}\right\}^{\beta}}-1}{\alpha-1}$$

The density functions $Z=(Y-\mu)/\sigma$ $(\text{for } z\in\mathbb{R})$ is specified as

$$f(z;\tau,\beta,\alpha,\mu,\sigma)=\frac{\tau\beta\log(\alpha)e^{z}e^{-e^{z}}\left(1-e^{-e^{z}}\right)^{\tau-1}e^{\frac{-\left(1-e^{-e^{z}}\right)^{\tau}}{1-\left(1-e^{-e^{z}}\right)^{\tau}}}}{\sigma(\alpha-1)\left[1-\left(1-e^{-e^{z}}\right)^{\tau}\right]^{2}}\left\{1-e^{\frac{-\left(1-e^{-e^{z}}\right)^{\tau}}{1-\left(1-e^{-e^{z}}\right)^{\tau}}}\right\}^{\beta-1}\alpha^{\left\{1-e^{\frac{-\left(1-e^{-e^{z}}\right)^{\tau}}{1-\left(1-e^{-e^{z}}\right)^{\tau}}}\right\}^{\beta}} \tag{19}$$

The standard LAPGOGEW density is Equation (19).

A regression model is constructed based on the LAPGOGEW for the response variable y_i with explanatory vector $\kappa_i^T=(\kappa_{i1},\kappa_{i2},\ldots,\kappa_{is})$ and parametric vector $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_s)^T$ as

$$y_i=\kappa_i^T\lambda+\sigma z_i,\quad i=1,\ldots n, \tag{20}$$

where $\mu_i=\kappa_i^T\lambda$ and z_i is the random error with density specified in Equation (19).

The density and survival functions of Y_i are specified as

$$f(y_i;\tau,\beta,\alpha,\sigma,\lambda^T)=\frac{\tau\beta\log(\alpha)e^{z_i}e^{-e^{z_i}}\left(1-e^{-e^{z_i}}\right)^{\tau-1}e^{\frac{-\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}}{\sigma(\alpha-1)\left[1-\left(1-e^{-e^{z_i}}\right)^{\tau}\right]^{2}}\left\{1-e^{\frac{-\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}\right\}^{\beta-1}$$
$$\times\alpha^{\left\{1-e^{\frac{-\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}\right\}^{\beta}}$$

and

$$S\left(y_i;\tau,\beta,\alpha,\sigma,\lambda^{T}\right)=1-\frac{\alpha^{\left\{1-e^{-\frac{\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}\right\}^{\beta}}-1}{\alpha-1}$$

where $z_i=\left(y_i-\kappa_i^{T}\lambda\right)/\sigma$. The group of breast cancer patients for which y_i is the log-lifetime or the non-informative log-censoring are represented by F and C with h and $n-h$ observations, respectively. Hence, the log-likelihood for $\Psi=\left(\tau,\beta,\alpha,\sigma,\lambda^{T}\right)^{T}$ is specified as

$$\ell(\Psi)=h\log\left(\frac{\tau\beta\log\alpha}{\sigma}\right)+\sum_{i\in F}z_i-\sum_{i\in F}e^{z_i}+(\tau-1)\sum_{i\in F}\log\left(1-e^{-e^{z_i}}\right)-\sum_{i\in F}\frac{\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}$$

$$+(\beta-1)\sum_{i\in F}\log\left[1-e^{\frac{\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}\right]+\sum_{i\in F}\left[1-e^{\frac{\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}\right]^{\beta}\log\alpha$$

$$\sum_{i\in C}\log\left(1-\frac{\alpha^{\left\{1-e^{-\frac{\left(1-e^{-e^{z_i}}\right)^{\tau}}{1-\left(1-e^{-e^{z_i}}\right)^{\tau}}}\right\}^{\beta}}-1}{\alpha-1}\right)\tag{21}$$

Maximizing Equation (21) numerically using R-programme (*optim function*), we obtain the MLE $\hat{\Psi}$.

RESULTS AND DISCUSSIONS

Simulation

Here, Monte Carlo simulations (MCS) with 1000 replications to inspect the estimation accuracy of the $\text{APGOGEW}(\tau,\beta,\eta,\varphi,\alpha)$ distribution with sample-sizes (n) = 50, 100, 150, 300, and 500 is executed. The samples are generated from Equation (8) and the true parameters (Pa.) utilized for initializing the MCS are: $\tau=1.1, \beta=1.5, \eta=0.3, \varphi=1.6, \alpha=1.3$. The MCS were executed using the R-programme.

The average estimates (AEs), Bias, mean square errors (MSEs) and average interval length (AILs) of the 95% confidence intervals (CIs) are reported in Table 2. The AEs tend to the true parameters and the MSEs decrease to small values when n increases, thus showing that the estimates are consistent.

Table 2. MCS findings for the APGOGEW.

n	Pa.	AE	Bias	MSE	Lower (95%)	Upper (95%)	AIL
50	τ	0.141	-0.959	0.920	0.135	0.148	0.014
	β	1.746	0.246	0.920	1.720	1.773	0.053
	η	0.151	-0.149	0.070	0.139	0.163	0.024
	φ	1.807	0.207	0.024	1.769	1.845	0.076
	α	1.498	0.198	0.062	1.448	1.548	0.100
100	τ	0.140	-0.960	0.922	0.136	0.144	0.009
	β	1.746	0.246	0.068	1.730	1.763	0.033
	η	0.147	-0.153	0.024	0.141	0.152	0.012
	φ	1.780	0.180	0.046	1.758	1.803	0.045
	α	1.469	0.169	0.060	1.434	1.504	0.070
200	τ	0.141	-0.959	0.920	0.138	0.144	0.005
	β	1.749	0.249	0.067	1.739	1.759	0.020
	η	0.149	-0.151	0.023	0.145	0.153	0.007
	φ	1.752	0.152	0.031	1.739	1.764	0.024
	α	1.451	0.151	0.048	1.429	1.473	0.044
300	τ	0.140	-0.960	0.922	0.138	0.142	0.004
	β	1.755	0.255	0.069	1.747	1.762	0.015
	η	0.150	-0.150	0.023	0.147	0.153	0.005
	φ	1.738	0.138	0.025	1.730	1.747	0.017

	α	1.451	0.151	0.043	1.435	1.467	0.032
500	τ	0.139	-0.961	0.924	0.138	0.141	0.003
	β	1.763	0.263	0.073	1.758	1.769	0.011
	η	0.151	-0.149	0.023	0.149	0.153	0.004
	φ	1.724	0.124	0.019	1.719	1.730	0.011
	α	1.450	0.150	0.035	1.440	1.460	0.020

The MCS values in Table 2 specify that the AEs converges to the true parameters of the APGOGEW distribution. The biases and MSEs tend to zero when n increases, which shows the consistency of the APGOGEW estimators. However, some of the parameter estimates are less accurate. Generally, the MCS findings suggest that larger sample sizes and the appropriate choice of selected parameter values are crucial for accurate parameter estimation of the APGOGEW distribution.

Applications

The fitness of the proposed models is demonstrated by means of two real-datasets applications using the R-programme (R Core, 2022). The selection of the model is based on the Akaike information criterion (AIC), Consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), Kolmogorov-Smirnov (KS) (its p-value), Cramer-von Mises (CVM) and Anderson-Darling (ADG) statistics. The model with stronger evidence of a good fit comes with least statistics value. Additionally, the graphical plots of the data histograms, estimated density and cumulative functions, and the empirical cumulative function are also vital in identifying the best fitting model.

Bladder Cancer Data

The first dataset comprises the remission times (in months) of 36 bladder cancer patients (Klakattawi, 2022). The observations are: 0.315,0.496, 0.616, 1.145, 1.208, 1.263, 1.414, 2.025, 2.036, 2.162, 2.211, 2.370, 2.532, 2.693, 2.805, 2.910, 2.912, 3.192, 3.263, 3.348, 3.348, 3.427, 3.499, 3.534, 3.767, 3.751, 3.858, 3.986, 4.049, 4.244, 4.323, 4.381, 4.392, 4.397, 4.647, 4.753, 4.929, 4.973, 5.074, 5.381. Some descriptive graphs for the data are depicted in Figure 4 and it is construed that this data is naturally leptokurtic and left-skewed.

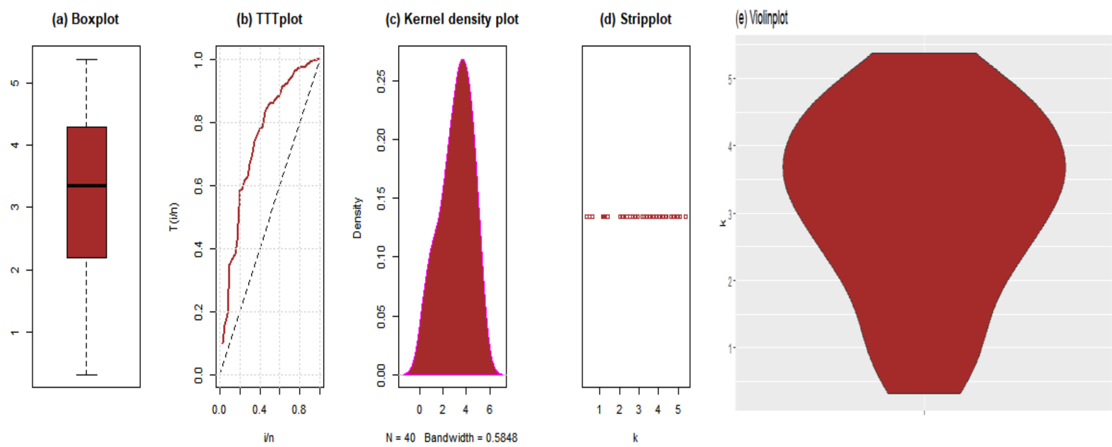

Figure 4. Box plot, TTT, kernel-density, strip, and Violin plots of the data.

The APGOGEW is compared with the alpha power Weibull (APW), generalized odd generalized exponentiated Weibull (GOGEW), exponentiated Weibull (EW), alpha power generalized exponentiated Weibull (APGEW) and Weibull (W) distributions. The EW and W distributions are familiar distributions. Table 3 reports the MLEs and standard errors (SEs) of the fitted models for the dataset. The empirical findings reveal that all models produce reliable estimates (with the exception of APGEW).

Table 3. Empirical results from fitted models to the cancer dataset

Model	MLE (SEs)				
APGOGEW$(\tau,\beta,\eta,\varphi,\alpha)$	22.416	0.139	1.894	1.575	14.499
	(0.0147)	(0.036)	(0.015)	(0.015)	(11.400)
APW(η,φ,α)	1.887	1.544	190.448		
	(0.203)	(0.183)	(124.575)		
GOGEW$(\tau,\beta,\eta,\varphi)$	17.569	0.267	1.967	1.465	
	(0.274)	(0.049)	(0.140)	(0.118)	
EW(η,φ,β)	1.820	1.651	6.010		
	(0.202)	(0.216)	(1.546)		
APGEW$(\tau,\beta,\eta,\varphi,\alpha)$	0.437	0.133	4.455	10.161	2.303
	(0.150)	(0.056)	(0.157)	(0.225)	(3.271)
W(η,φ)	3.141	0.869			
	(0.533)	(0.137)			

Based on the findings presented in Table 4, which specifically contains models with correct estimates, it can be observed that the APGOGEW model exhibits the lowest values for the adequacy measures, specifically the KS, CVM, and ADG. This suggests that the APGOGEW model offers the most accurate match to the cancer data compared to the other models that were examined. The GOGEW model is ranked second based on these measures. Conversely, the EW and W models have the greatest values for CVM and ADG, as well as other metrics, signifying a relatively inadequate fit to the data.

Figure 5 indicates that the APGOGEW model's PDF and CDF are more similar to the histogram and empirical CDF of the data. Based on these findings, the APGOGEW model is recommended as the best fitting model.

Table 4. Adequacy measures for the models.

Model	CVM	ADG	AIC	BIC	CAIC	HQIC	KS	p-value
APGOGEW	0.062	0.416	130.7	139.1	132.4	133.7	0.14	0.40
APW	0.208	1.294	139.4	144.5	140.1	141.3	0.25	0.01
GOGEW	0.152	0.960	140.8	147.6	142.0	143.3	0.22	0.04
EW	0.315	1.894	144.4	149.5	145.1	146.2	0.29	0.003
W	0.247	1.518	174.7	178.1	175.0	175.9	0.25	0.01

Table 5 presents the likelihood ratio (LR) tests for the GOGEW, APW, and EW models, which are all special cases of the APGOGEW model. In all three situations, the null

hypotheses are rejected with low p-values, showing that the APGOGEW model fits the data much better than the other three models. This shows that the APGOGEW model is a better fit for this dataset.

Table 5. LR tests results for the cancer dataset

Model	Hypotheses	LR	p-values
APGOGEW vs GOGEW	$\mathrm{H}_0 : \alpha = 1$ **vs** $\mathrm{H}_1 : \mathrm{H}_0$ **is false**	12.17	0.0001
APGOGEW vs APW	$\mathrm{H}_0 : \tau = \beta = 1$ **vs** $\mathrm{H}_1 : \mathrm{H}_0$ **is false**	13.00	0.0001
APGOGEW vs EW	$\mathrm{H}_0 : \tau = \alpha = 1$ **vs** $\mathrm{H}_1 : \mathrm{H}_0$ **is false**	18.00	<0.0001

The plots of the profile log-likelihood function versus some parameter values (with fixed MLEs of other parameters) for the first data are displayed in Figures 6. We can see the approximate intervals for each parameter that maximize the profile log-likelihood function. However, there are evidence of a monotone log-likelihood in Figures 6(a) and 6(e) given that the functions do not have the maximum in the range taken for τ and α.

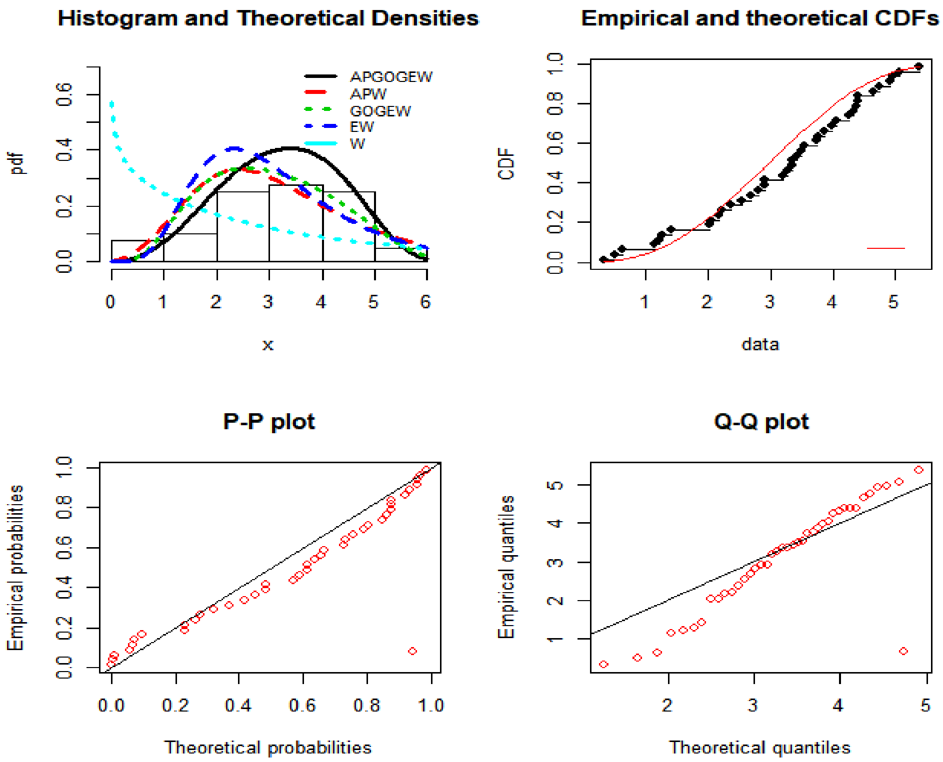

Figure 5. Estimated PDFs and APGOGEW CDF, P-P and Q-Q plots for Cancer data.

Regression for Breast Cancer Data

The data comprises the lifetime (in months) of 54 individuals diagnosed with Breast cancer via DBT screening in ATBU teaching hospital (ATBUTH)-Bauchi, Nigeria. The data were collected from the cancer record unit of the ATBUTH and will be made available upon request. The response variable y_t denote the survival time of the breast cancer patients at the facility.

Under right censoring (0 = censored, 1 = observed lifetime): the observations $(i = 1,\ldots,54)$ percentage of censoring is around 58.52%. The explanatory variables considered κ_{i1} : age (in years), κ_{i2} : clinical-stage (I, II, III and IV), and κ_{i3} : BMI (kg/m^2). The proposed regression model is

$$y_i = \lambda_0 + \lambda_1\kappa_{i1} + \lambda_2\kappa_{i2} + \lambda_3\kappa_{i3} + \sigma z_i, \quad i = 1,\ldots,54,$$

where z_i follows the density in Equation (19). Table 6 reports the MLEs with SEs in parentheses, and p-values in brackets for the regression models. The LAPGOGEW regression results are compared to the log-GOGEW (LGOGEW), log-APW (LAPW) and log-EW (LEW) regressions. The estimates reported in Table 6 indicates that the explanatory variables; age, clinical stage and BMI are significant at $\alpha = 0.05$ level of significance. The negative signs of λ_2 and λ_3 means that the higher the clinical stage and BMI, shorter the time to failure.

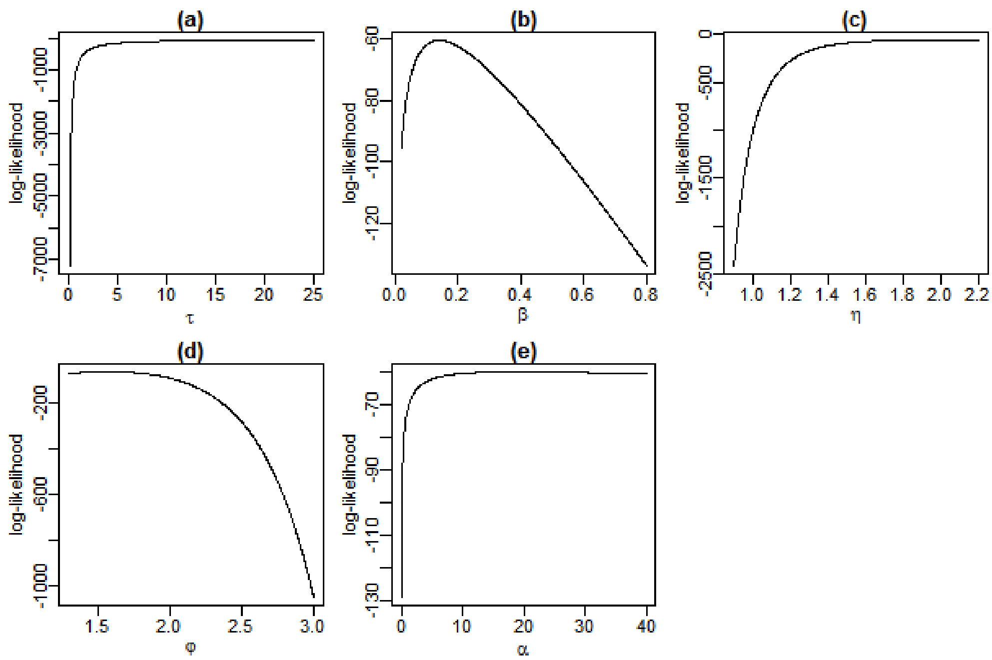

Figure 6. Profile log-likelihood functions for first data (a–e).

Table 6. The MLEs, SEs and p-values for the regression models.

Model	α	τ	β	σ	λ_0	λ_1	λ_2	λ_3
LAPGOGEW	15.585 (2.573)	0.944 (0.197)	0.055 (0.020)	0.282 (0.024)	20.129 (0.207)	0.045 (0.018) [0.0172]	-3.024 (0.189) [<0.0001]	-0.220 (0.017) [<0.0001]
LGOGEW	-	0.909 (0.473)	0.181 (0.078)	0.053 (0.063)	6.664 (0.755)	-0.010 (0.002) [<0.0001]	-0.332 (0.133) [0.016]	-0.040 (0.029) [0.171]

LAPW	6.291 (5.846)	-	-	0.327 (0.130)	4.466 (0.769)	-0.007 (0.010) [0.492]	-0.367 (0.081) [<0.0001]	0.024 (0.040) [0.546]
LEW	-	-	4.243 (2.908)	0.553 (0.297)	4.470 (1.015)	-0.015 (0.010) [0.147]	-0.391 (0.082) [<0.0001]	0.030 (0.046) [0.514]

More so, the LAPGOGEW regression has the least criterion values as reported in Table 7. For the residual analysis of fitted LAPGOGEW regression, the quantile residuals (qrs) introduced by Dunn and Smyth (1996) is adopted.

$$\mathrm{qr}_i = \Phi^{-1}\left(\frac{\hat{\alpha}^{\left\{ 1 - e^{ \frac{-\left(1-e^{-e^{\hat{z}_i}}\right)^{\hat{\tau}}}{1-\left(1-e^{-e^{\hat{z}_i}}\right)^{\hat{\tau}}} } \right\}^{\hat{\beta}}} - 1}{\hat{\alpha} - 1} \right),$$

where $\hat{z}_i = \left(y_i - \boldsymbol{\kappa}_i^{T}\hat{\boldsymbol{\lambda}}\right)/\hat{\sigma}$ and $\Phi(\cdot)^{-1}$ is the standard normal quantile function. As observed in Figure 7(a), no observation is outside the range [-1.5,1.5], this indicates that the qrs are distributed randomly and the normal probability plot depicted in Figure 7(b) indicates that the qrs follows the standard normal distribution approximately. Hence, the evidence supports the LAPGOGEW regression assumptions.

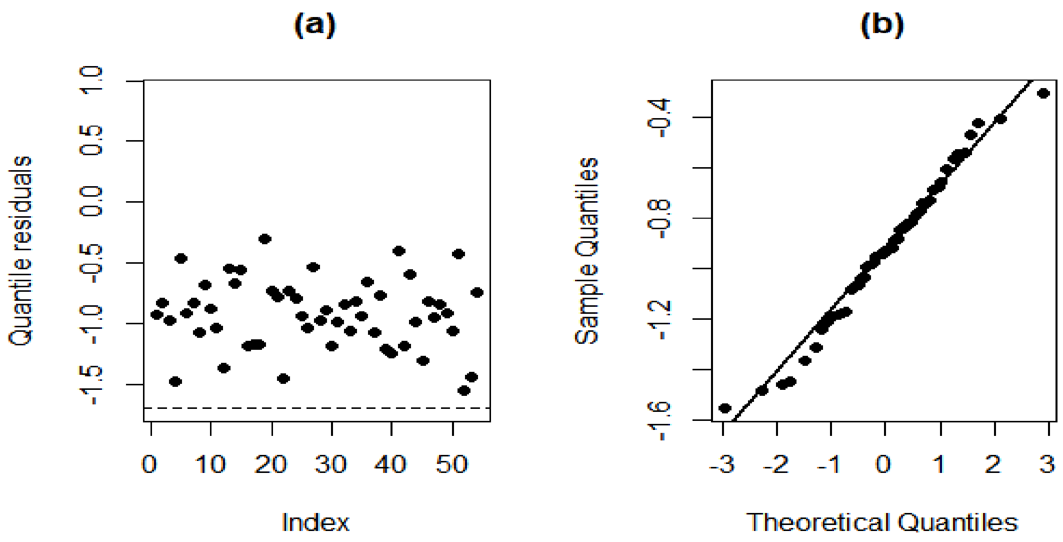

Figure 7. (a) Index plot and (b) Normal probability plot for the breast cancer data.

Table 7. Criteria for the fitted regression models.

Model	AIC	AICc	BIC	HQIC
LAPGOGEW	-47.10	-41.98	-31.19	-40.96
LGOGEW	57.05	61.14	70.97	62.42
LAPW	53.49	56.69	65.42	58.09
LEW	50.73	53.93	62.67	55.34

CONCLUSION

We presented a new five-parameter alpha power generalized odd generalized exponentiated Weibull (APGOGEW) model, which comprise special cases like the generalized odd generalized exponentiated Weibull, alpha power Weibull, exponentiated Weibull and Weibull distributions. Some of its structural properties were derived. The simulation study and empirical result from the real-data application disclosed the consistency of the maximum likelihood estimators and fitness of the APGOGEW model. More so, the log-APGOGEW regression model is constructed for censored data and its applicability to breast cancer data is demonstrated. The proposed model offered the best fit when compared to other well-known models.

REFERENCES

Abdulkadir, A., Adubisi, O. D. & Madaki, R. (2024). A new family of distributions based on amalgamation of two methods with an application to the Rayleigh model. European Journal of Statistics, 4(2024), 1-19

Adamu, P. I., Adamu, M. O., Okagbue, H. I., Opoola L. & Bishop S. A. (2019). Survival analysis of cancer patients in North Eastern Nigeria from 2004 – 2017 – A Kaplan - Meier Method. Open Access Macedonian Journal of Medical Sciences, 7(4), 642-649.

American Cancer Society [ACS] (2019). Breast cancer facts & figures 2019-2020. Atlanta: American Cancer Society, Inc. p.1

Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5, 236–244.

Feleke, B., Tesfaw, L. M. & Mitku, A. A. (2022). Survival analysis of women breast cancer patients in Northwest Amhara, Ethiopia. Front. Oncol. 12: 1041245. for Statistical Computing, Vienna, Austria.

Klakattawi, H. S. (2022). Survival analysis of cancer patients using a new extended Weibull distribution. PLoS ONE, 17(2), e0264229.

Mahdavi, A. & Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. Commun. Stat. Theory Methods, 46, 6543-6557.

Marinho, P. R. D., Silva, R. B., Bourguignon, M., Cordeiro, G. M., & Nadarajah, S. (2019). AdequacyModel: An R package for probability distributions and general-purpose optimization. PLoS ONE, 14, e0221487.

Misganaw, M., Zeleke, H., Mulugeta, H. & Assefa, B. (2023). Mortality rate and predictors among patients with breast cancer at a referral hospital in northwest Ethiopia: A retrospective follow-up study. PLoS ONE 18(1), e0279656

Nabegu, A. A., Muhammad, M. U. & Abdullahi, N. (2023). Survival analysis of breast cancer patients in Kano state. FUDMA Journal of Sciences (FJS), 7(2), 199 – 205.

Nnaka, J. A., Likkason, A., Prudence, K. M. & Ojima-Ojo, E. V. (2022). Age distribution and histo-epidemiological profile of breast cancer cases in Bauchi, North-eastern Nigeria: A five-year review. Current Medical Research, 1(1), 1-7.

R Core Team. (2022). A Language and Environment for Statistical Computing. R Foundation

Weibull, W. (1951). A statistical distribution function of wide applicability. Journal of applied mechanics. 18(3), 293–297. https://doi.org/10.1115/1.4010337