TWMS J. App. and Eng. Math. V.16, N.9, 2026, pp.1057-1077

PROPERTIES AND REAL DATA APPLICATION OF THE SINE ALPHA POWER BURR XII REGRESSION MODEL

AHMED MAHDI SALIH[1*], AMAL SADEQ HAMOODI[2]

Abstract. We propose the Sine Alpha Power Burr XII (SAP Burr XII) distribution, obtained by applying the Sine Alpha Power transformation to the Burr XII distribution. and develop an accelerated failure time (AFT) regression model, with parameters estimated via maximum likelihood. A simulation study across various sample sizes confirms the consistency of the estimators. Additionally, model diagnostics based on Cox–Snell, deviance, and randomized quantile residuals are examined. Application to a real lung cancer dataset demonstrates that the SAP Burr XII model outperforms classical AFT models.

Keywords: Burr XII distribution, Sine Alpha Power transformation, SAP-Burr XII distribution, Regression model, Accelerated failure time.

AMS Subject Classification: 62J05, 62P10.

1. Introduction

Applied statistics, especially in survival analysis, reliability engineering, medicine and finance, frequently deal with data that have complex features not captured by classical distributions. They are characterized by asymmetry (skewness), heavy tails, multimodality and non-monotonic hazard rates including bathtub or unimodal hazards. Burr [9] ,1942 proposed a distribution, now known as the Burr XII, that has been popular for many years as a flexible model for the shape of hazard functions. The standard Burr XII distribution, however, does not suit some situations with extremely heavy tails and some types of asymmetry which can often occur in real datasets. To overcome these shortcomings, a number of generalizations have been suggested like exponentiated family, beta generated family, Kumaraswamy generated family and transformed transformed families. The Sine Alpha Power (SAP) family of recent transformation techniques Souza et al. [24], 2022 is a powerful and flexible technique for generating new distributions that have improved ability to model complex data structures. The Sine Alpha Power Burr XII (SAP-Burr XII)

[1*]Department of Statistics, College of Administration and Economics, Wasit University, Kut, Iraq.
amahdi@uowasit.edu.iq; 0000-0002-6109-224X.
[2] Mathematics Department, College of Education ,Mustansiriya University, Iraq.
d.amalsadik@uomustansiriyah.edu.iq;0000-0001-7496-7933
[*] Corresponding author.
§ Manuscript received: August 15, 2025; accepted: February 17, 2026.
TWMS Journal of Applied and Engineering Mathematics, Vol.16, No.9;

distribution can be derived by multiplying the SAP transformation with the Burr XII distribution and is expected to have improved properties in modeling real data with the shape of the hazards to be nonlinear. Most real-world research questions have to do with understanding the relationship between a response variable and one or more explanatory variables (covariates) that affect it. For instance, in the medical area, patient survival times can be expected to be modeled as a function of age, treatment, and genetic markers. For engineering purposes, the time to failure of a component can be dependent upon the temperature, pressure and quality of the material. Covariate effects can be added to the distribution of the response variable in a systematic manner using regression models. For survival/failure time data, the normality and constant variance assumptions of standard linear regression do not necessarily hold. However, flexible distribution parametric regression models like accelerated failure time (AFT) models provide a more realistic approach. The highly flexible distribution regression model can provide accurate estimation of covariate effects with the consideration of non-normality, heavy tails, and non-monotonic hazard functions. This is why it is crucial to extend a flexible distribution to a regression context rather than simply extending it.

Distribution Theory and regression modeling are closely related and natural. Though flexible [23], a distribution is still merely a description that does not provide any information when it comes to covariates. On the other hand, a regression model based on a rigid distribution can lead to biased estimates of model parameters, wrong standard errors and false conclusions. In this research, the proposed SAP-Burr XII distribution is inserted into a regression framework, which has two complementary objectives. Firstly, the marginal distribution of the response is realistic and can accommodate complex features in the data. Second, a robust estimation of the effects of covariates can be achieved efficiently and accurately. When there is no flexible baseline distribution, model misspecification occurs since the regression model cannot adequately represent unobserved heterogeneity. In the absence of the regression structure, the flexible distribution will not be able to make an answer to causal or associational research questions. With this in mind, the present study fills this void by introducing a parametric regression model under the SAP-Burr XII distribution, which is a combination of distributional flexibility and interpretability of covariate effects.

The Burr XII distribution was originally introduced by Burr (1942) as one of twelve cumulative distribution functions, and Tadikamalla [25] 1980 later demonstrated its flexibility in approximating gamma, Weibull, and log-logistic distributions. Zimmer et al. [28] 1998 showed that Burr XII outperforms the Weibull distribution in engineering failure data, particularly when the hazard rate is non-monotonic. Alizadeh et al. [4] 2017 proposed the odd log-logistic Burr XII distribution and studied its theoretical properties. In the context of regression models, Alizadeh et al. [3] 2020 developed the odd log-logistic Burr XII regression model for survival data and applied it to cancer patient data, showing better fit than standard Burr XII and Weibull regression models. Bantan et al. [7] 2021 introduced the power Burr XII regression model with applications to COVID-19 data. Cordeiro et al. [10] 2020 proposed a log-gamma generalized Weibull regression model for censored data, while Korkmaz and Genç [18] 2021 developed the log-odd exponential Weibull regression model for lifetime data. Emiliano et al. [14] 2022 proposed a Burr XII regression model with varying dispersion for counting data. Bhatti et al. [8] 2023 developed the Gumbel Burr XII regression model for asymmetric heavy-tailed data. Rasekhi et al. [22] 2024 developed a regression model based on the gamma exponentiated exponential Weibull distribution, while Hamedani et al. [15] 2024 introduced a beta Burr XII regression model for credit risk. Regarding the Sine Alpha Power family, Souza et al. [24]

2022 formally introduced the SAP family. Despite these extensive contributions, none of these studies have combined the Sine Alpha Power transformation with the Burr XII distribution within a regression framework, confirming the novelty of the present research.

After a thorough and systematic review of the literature from 2020 to 2026, several critical and interconnected gaps have been identified [1],[5],[13]. First, from a theoretical perspective, no previous study has mathematically derived the Sine Alpha Power Burr XII (SAP-Burr XII) distribution. Consequently, its fundamental statistical properties have never been theoretically established. Second, from a methodological standpoint, no regression model has yet been formulated using the SAP-Burr XII distribution. Third, from a computational perspective, there is no available software implementation for fitting the SAP-Burr XII regression model. Fourth, from an empirical perspective, the SAP-Burr XII regression model has never been applied to any real dataset. This research aims to fill this gap entirely.

The remainder of this paper is organized as follows. Section 2 presents the proposed SAP-Burr XII distribution and its properties. Section 3 introduces the SAP-Burr XII regression model. Section 4 presents Maximum Likelihood estimation. Section 5 present a simulation study. Section 6 provides a real data application. Section 7 offers concluding remarks.

2. Sine Alpha Power Burr XII Distribution (SAP-Burr XII)

2.1. **Definition of the Baseline Burr XII Distribution.** Let T be a non-negative continuous random variable. The Burr XII distribution with scale parameter $c > 0$ and shape parameters $\alpha > 0$ and $k > 0$ is defined by its cumulative distribution function (CDF) and probability density function (PDF) as follows [27]:

$$F(t) = 1 - \left[1 + \left(\frac{t}{c}\right)^{\alpha}\right]^{-k}, \qquad t > 0. \tag{1}$$

$$f(t) = \frac{k\alpha}{c}\left(\frac{t}{c}\right)^{\alpha-1}\left[1 + \left(\frac{t}{c}\right)^{\alpha}\right]^{-k-1}, \qquad t > 0. \tag{2}$$

The Burr XII distribution is a flexible lifetime distribution that includes the log-logistic distribution as a special case when $k = 1$. Its hazard function can be decreasing, increasing, or unimodal, making it suitable for a wide range of applications [19].

2.2. **Definition of the Sine Alpha Power Transformation.** Let $F_0(t)$ be any valid cumulative distribution function (CDF). The Sine Alpha Power (SAP) transformation defines a new CDF as

$$F(t) = \sin\left(\frac{\pi}{2}\left[F_0(t)\right]^{\theta}\right), \qquad \theta > 0, \tag{3}$$

where $\theta > 0$ is an additional shape parameter. The transformation is called the *Sine Alpha Power (SAP)* transformation because it combines the sine function with a power transformation of the baseline CDF.

Theoretical Note. The SAP transformation satisfies all the properties of a valid cumulative distribution function: it is non-decreasing, right-continuous, and satisfies [19]

$$\lim_{t\to-\infty} F(t) = 0, \qquad \lim_{t\to\infty} F(t) = 1.$$

The parameter θ provides additional flexibility, allowing the model to accommodate data with a wide variety of distributional characteristics and complex shapes of the hazard rate.

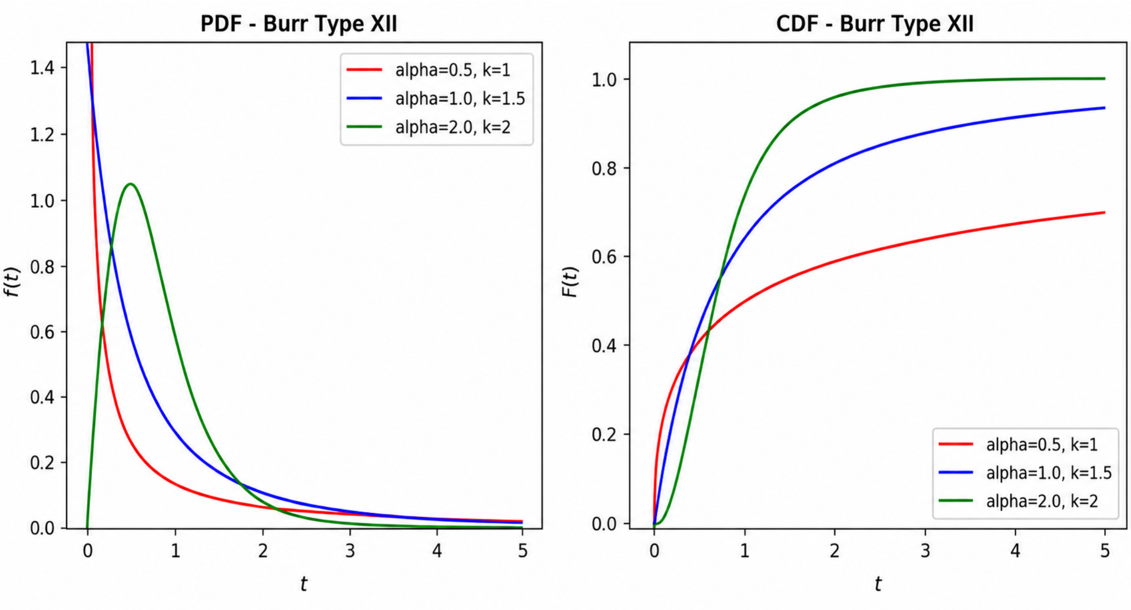

FIGURE 1. PDF and CDF for Burr type XII at different values of parameters

2.3. **Definition of the SAP-Burr XII Distribution.** Applying the Sine Alpha Power (SAP) transformation to the Burr XII cumulative distribution function (CDF), we obtain the *Sine Alpha Power Burr XII* (SAP–Burr XII) distribution. Its cumulative distribution function is given by [20]

$$F(t) = \sin\left(\frac{\pi}{2}\left[1-\left(1+\left(\frac{t}{c}\right)^{\alpha}\right)^{-k}\right]^{\theta}\right), \qquad t>0. \tag{4}$$

where $c>0$ is the scale parameter, $\alpha>0$, $k>0$, and $\theta>0$ are shape parameters. The SAP–Burr XII distribution reduces to the standard Burr XII distribution as a limiting case when $\theta=1$ and the argument of the sine function is small. Owing to its four parameters, the SAP–Burr XII distribution provides considerable flexibility for modeling data with a variety of shapes and hazard rate behaviors.

2.4. **Probability Density Function (PDF).** The probability density function (PDF) of a continuous random variable is obtained by differentiating its cumulative distribution function (CDF). For the SAP–Burr XII distribution, differentiating $F(t)$ with respect to t yields

$$\begin{aligned} f(t) = \frac{\pi\alpha\theta k}{2c}\left(\frac{t}{c}\right)^{\alpha-1}\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k-1}\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta-1} \\ \times\cos\left(\frac{\pi}{2}\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta}\right), \qquad t>0. \end{aligned} \tag{5}$$

The probability density function exists for all $t>0$ and satisfies the normalization condition

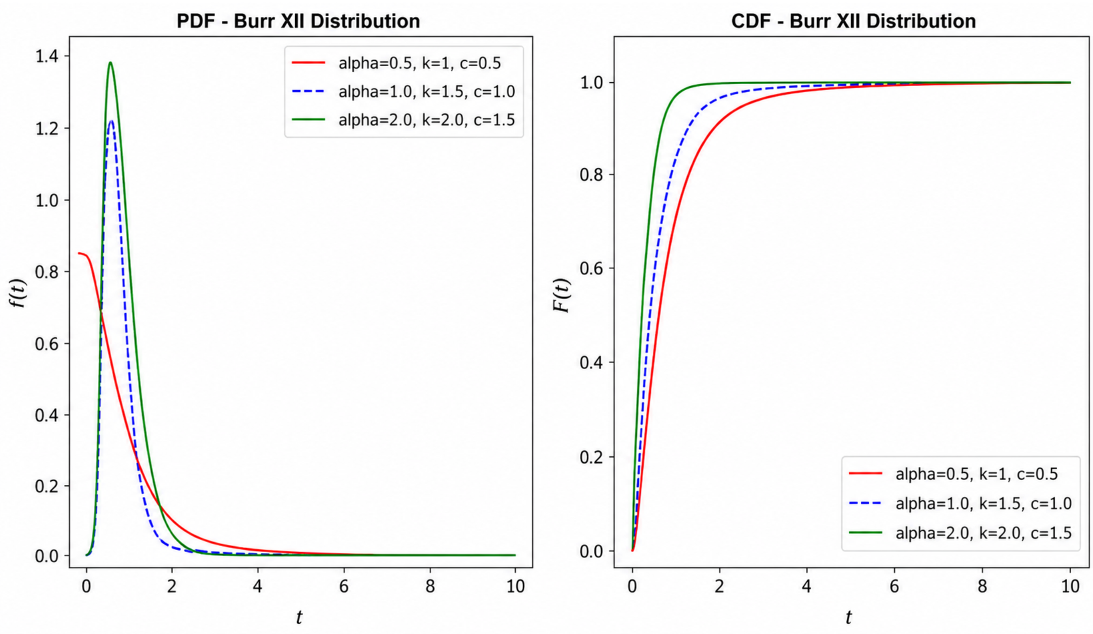

FIGURE 2. PDF and CDF for SAP-Burr type XII at different values of parameters

$$\int_0^\infty f(t)\,dt = 1.$$

Furthermore, the presence of the cosine term ensures that the probability density function remains non-negative over its support, provided that the argument of the cosine function lies within the interval $[0, \pi/2]$. Consequently, the SAP–Burr XII distribution defines a valid probability density function [21].

The survival function is defined as the complement of the cumulative distribution function, that is,

$$S(t) = 1 - F(t), \qquad t > 0. \tag{6}$$

The survival function is monotonically decreasing, satisfying

$$S(0) = 1, \qquad \lim_{t\to\infty} S(t) = 0.$$

It is widely used in survival analysis and reliability theory to quantify the probability that a unit or individual survives beyond a specified time t.

2.5. **Hazard Rate Function.** The hazard rate function, $h(t)$, also known as the *instantaneous failure rate*, is defined as [17]

$$h(t) = \frac{f(t)}{S(t)}, \qquad t > 0, \tag{7}$$

where $f(t)$ and $S(t)$ denote the probability density function and the survival function, respectively. Equivalently,

$$h(t) = \lim_{\Delta t\to 0} \frac{\Pr(t \le T < t + \Delta t \,|\, T \ge t)}{\Delta t}. \tag{8}$$

For the SAP–Burr XII distribution, the hazard rate function is given by

$$h(t)=\frac{\frac{\pi\alpha\theta k}{2c}\left(\frac{t}{c}\right)^{\alpha-1}\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k-1}\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta-1}\cos\left(\frac{\pi}{2}\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta}\right)}{1-\sin\left(\frac{\pi}{2}\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta}\right)}, t>0. \tag{9}$$

The hazard rate function of the SAP–Burr XII distribution is highly flexible and is capable of exhibiting a variety of shapes, including increasing (increasing failure rate, IFR), decreasing (decreasing failure rate, DFR), bathtub-shaped (initially decreasing and then increasing), and unimodal (initially increasing and then decreasing). This flexibility makes the SAP–Burr XII distribution particularly suitable for modeling a wide range of survival and reliability data [27].

2.6. **Quantile Function.** The quantile function $Q(u)$, for $u \in (0,1)$, is defined as the inverse of the cumulative distribution function (CDF), that is [16],

$$Q(u)=F^{-1}(u), \qquad 0<u<1. \tag{10}$$

It represents the value of t such that

$$\Pr(T \le t)=u.$$

To derive the quantile function of the SAP–Burr XII distribution, we solve $F(t)=u$. Thus,

$$u=\sin\left(\frac{\pi}{2}\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta}\right), \tag{11}$$

$$\frac{2}{\pi}\sin^{-1}(u)=\left\{1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}\right\}^{\theta}, \tag{12}$$

$$1-\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}=\left(\frac{2}{\pi}\sin^{-1}(u)\right)^{1/\theta}, \tag{13}$$

$$\left[1+\left(\frac{t}{c}\right)^{\alpha}\right]^{-k}=1-\left(\frac{2}{\pi}\sin^{-1}(u)\right)^{1/\theta}, \tag{14}$$

$$1+\left(\frac{t}{c}\right)^{\alpha}=\left[1-\left(\frac{2}{\pi}\sin^{-1}(u)\right)^{1/\theta}\right]^{-1/k}. \tag{15}$$

Hence, the quantile function is

$$Q(u)=c\left\{\left[1-\left(\frac{2}{\pi}\sin^{-1}(u)\right)^{1/\theta}\right]^{-1/k}-1\right\}^{1/\alpha}, \qquad 0<u<1. \tag{16}$$

The quantile function is particularly useful for generating random samples from the SAP–Burr XII distribution using the inverse transform sampling method. The median is obtained by setting $u=0.5$, yielding

$$\text{Median} = Q(0.5) = c\left\{\left[1-\left(\frac{2}{\pi}\sin^{-1}(0.5)\right)^{1/\theta}\right]^{-1/k}-1\right\}^{1/\alpha}. \tag{17}$$

3. SAP-Burr XII Regression Model

3.1. **Model Specification.** When using a survival analysis model, the accelerated failure time (AFT) framework assumes that the effect of covariates is to accelerate or decelerate the time to failure. For the SAP–Burr XII distribution, the covariates are incorporated through the scale parameter c_i using a log-link function [16],

$$c_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) = \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right), \tag{18}$$

where

$$\mathbf{x}_i^\top = (1, x_{i1}, x_{i2}, \ldots, x_{ip})$$

is the vector of covariates for the i-th individual, and

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)^\top$$

is the vector of regression coefficients. Under this formulation, the conditional distribution of the survival time T_i, given the covariate vector $\mathbf{x}_i$, follows a SAP–Burr XII distribution, namely,

$$T_i \mid \mathbf{x}_i \sim \text{SAP-Burr XII}\,(c_i, \alpha, k, \theta), \tag{19}$$

where

$$c_i = \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right),$$

while $\alpha > 0$, $k > 0$, and $\theta > 0$ are the shape parameters of the distribution.

4. Maximum Likelihood Estimation

For observed survival times $t_1, t_2, \ldots, t_n$, assuming complete and uncensored observations, let

$$\boldsymbol{\psi} = (\alpha, k, \theta, \beta_0, \beta_1, \ldots, \beta_p)^\top$$

denote the vector of unknown parameters. Under the SAP–Burr XII accelerated failure time model, the scale parameter for the i-th individual is

$$c_i = \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right).$$

Define

$$A_i = 1 + \left(\frac{t_i}{c_i}\right)^{\alpha}, \qquad B_i = 1 - A_i^{-k},$$

and

$$Z_i = \frac{\pi}{2} B_i^{\theta}.$$

The probability density function of T_i, conditional on $\mathbf{x}_i$, is

$$f(t_i \mid \mathbf{x}_i) = \frac{\pi\alpha\theta k}{2c_i}\left(\frac{t_i}{c_i}\right)^{\alpha-1} A_i^{-k-1} B_i^{\theta-1}\cos(Z_i), \qquad t_i > 0. \tag{20}$$

Hence, the likelihood function is [6]

$$L(\boldsymbol{\psi}) = \prod_{i=1}^{n}\left[\frac{\pi\alpha\theta k}{2c_i}\left(\frac{t_i}{c_i}\right)^{\alpha-1} A_i^{-k-1} B_i^{\theta-1}\cos(Z_i)\right]. \tag{21}$$

Taking the natural logarithm gives the log-likelihood function

$$\begin{aligned}\ell(\boldsymbol{\psi}) &= \sum_{i=1}^{n}\log f(t_i \mid \mathbf{x}_i)\\ &= n\log\left(\frac{\pi}{2}\right) + n\log\alpha + n\log\theta + n\log k - \sum_{i=1}^{n}\log c_i\\ &\quad + (\alpha-1)\sum_{i=1}^{n}\log\left(\frac{t_i}{c_i}\right) - (k+1)\sum_{i=1}^{n}\log A_i\\ &\quad + (\theta-1)\sum_{i=1}^{n}\log B_i + \sum_{i=1}^{n}\log[\cos(Z_i)]\,. \end{aligned} \tag{22}$$

For convenience, write

$$\ell(\boldsymbol{\psi}) = K_1 + K_2 + K_3, \tag{23}$$

where

$$K_1 = n\log\left(\frac{\pi}{2}\right) + n\log\alpha + n\log\theta + n\log k - \sum_{i=1}^{n}\log c_i, \tag{24}$$

$$K_2 = (\alpha-1)\sum_{i=1}^{n}\log\left(\frac{t_i}{c_i}\right) - (k+1)\sum_{i=1}^{n}\log A_i + (\theta-1)\sum_{i=1}^{n}\log B_i, \tag{25}$$

and

$$K_3 = \sum_{i=1}^{n}\log[\cos(Z_i)]\,. \tag{26}$$

The score equations are obtained by differentiating the log-likelihood with respect to each unknown parameter,

$$\frac{\partial\ell(\boldsymbol{\psi})}{\partial\theta} = \frac{\partial K_1}{\partial\theta} + \frac{\partial K_2}{\partial\theta} + \frac{\partial K_3}{\partial\theta}, \tag{27}$$

$$\frac{\partial\ell(\boldsymbol{\psi})}{\partial\alpha} = \frac{\partial K_1}{\partial\alpha} + \frac{\partial K_2}{\partial\alpha} + \frac{\partial K_3}{\partial\alpha}, \tag{28}$$

$$\frac{\partial\ell(\boldsymbol{\psi})}{\partial k} = \frac{\partial K_1}{\partial k} + \frac{\partial K_2}{\partial k} + \frac{\partial K_3}{\partial k}, \tag{29}$$

and

$$\frac{\partial\ell(\boldsymbol{\psi})}{\partial\beta_j} = \frac{\partial K_1}{\partial\beta_j} + \frac{\partial K_2}{\partial\beta_j} + \frac{\partial K_3}{\partial\beta_j}, \qquad j = 0, 1, \ldots, p. \tag{30}$$

The maximum likelihood estimates are obtained by solving the nonlinear score equations

$$\frac{\partial \ell(\boldsymbol{\psi})}{\partial \theta} = 0, \qquad \frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha} = 0, \qquad \frac{\partial \ell(\boldsymbol{\psi})}{\partial k} = 0, \qquad \frac{\partial \ell(\boldsymbol{\psi})}{\partial \beta_j} = 0, \quad j = 0, 1, \ldots, p.$$

Since these equations have no closed-form solution, the parameter estimates are obtained numerically using iterative optimization algorithms such as the Newton–Raphson or BFGS methods. The likelihood presented in this study is derived under the assumption of complete (uncensored) observations. Although the real lung cancer dataset contains approximately 31.6% right-censored observations, the present implementation focuses on complete-data likelihood estimation. Extending the proposed SAP–Burr XII regression model to incorporate right-censored likelihoods will be considered in future research.

5. Simulation Study

5.1. **Objectives of the Simulation Study.** The main objectives of this simulation study are as follows [26]:

(1) To evaluate the performance of the maximum likelihood estimators (MLEs) of the parameters in the SAP–Burr XII regression model.
(2) To assess the finite-sample properties of the estimators in terms of their bias and mean squared error (MSE).
(3) To examine the coverage probabilities of the corresponding confidence intervals.
(4) To investigate the behavior of the estimators under different sample sizes.

5.2. **Simulation Design.** The simulation study was conducted under the following settings:

- **True parameter values:**

$$\alpha = 1.5, \qquad k = 2.0, \qquad \theta = 0.8, \qquad \beta_0 = 0.5, \qquad \beta_1 = -0.3.$$

- **Covariate generation:**

$$x_i \sim N(0,1), \qquad i = 1, 2, \ldots, n.$$

- **Sample sizes:**

$$n \in \{50,\ 100,\ 200,\ 500\}.$$

- **Number of Monte Carlo replications:**

$$R = 1000.$$

5.3. **Theoretical Steps of the Simulation.** The Monte Carlo simulation procedure is summarized in the following steps [2].

(1) Set the true parameter values

$$\alpha = 1.5, \quad k = 2.0, \quad \theta = 0.8, \quad \beta_0 = 0.5, \quad \beta_1 = -0.3.$$

(2) Specify the sample size n and the number of Monte Carlo replications

$$R = 1000.$$

(3) Initialize storage vectors (or matrices) to save the parameter estimates, standard errors, and evaluation measures.
(4) For each replication $r = 1, 2, \ldots, R$, perform the following steps:
   (a) Generate the covariates

$$x_i \sim N(0,1), \qquad i = 1, 2, \ldots, n.$$

(b) Compute the scale parameter for each observation using the AFT model

$$c_i = \exp(0.5 - 0.3x_i).$$

(c) Generate random variables

$$u_i \sim U(0,1), \qquad i = 1, 2, \ldots, n.$$

(d) Generate survival times from the SAP–Burr XII distribution using the inverse transform method

$$t_i = c_i \left\{ \left[1 - \left(\frac{2}{\pi} \sin^{-1}(u_i) \right)^{1/\theta} \right]^{-1/k} - 1 \right\}^{1/\alpha}.$$

(e) Estimate the model parameters using the maximum likelihood method with the BFGS optimization algorithm.

(f) Compute the standard errors from the inverse Fisher information matrix.

(g) Store the parameter estimates and their corresponding standard errors.

(5) After completing all replications, compute the evaluation measures:
- Average Estimate (AE),
- Bias,
- Relative Bias (RB),
- Mean Squared Error (MSE),
- Root Mean Squared Error (RMSE),
- Coverage Probability (CP).

(6) Repeat Steps 2–5 for each sample size

$$n \in \{50, 100, 200, 500\}.$$

TABLE 1. Simulation results for the SAP–Burr XII regression model $n = 50$.

Parameter	True	Average Estimate	Bias	RB (%)	MSE	RMSE	CP (%)
α	1.5	1.4872	-0.0128	-0.85	0.0452	0.2126	93.2
k	2.0	1.9785	-0.0215	-1.07	0.0678	0.2604	92.8
θ	0.8	0.7912	-0.0088	-1.10	0.0189	0.1375	93.5
β_0	0.5	0.4934	-0.0066	-1.32	0.0234	0.1530	94.1
β_1	-0.3	-0.2956	0.0044	-1.47	0.0123	0.1109	94.3

5.4. **Simulation Results for Parameter Estimation.** The maximum likelihood estimators perform reasonably well for the smallest sample size ($n = 50$) with relatively small biases ranging from -1.47% to -0.85% in terms of relative bias. The highest mean squared error (MSE) is observed for the shape parameter k (0.0678), whereas the lowest MSE is obtained for the regression coefficient β_1 (0.0123). The empirical coverage probabilities (CPs) of the 95% confidence intervals range from 92.8% to 94.3%, which are slightly below the nominal confidence level because of the asymptotic approximation used in the maximum likelihood method. These findings indicate that the estimators can still be used with small samples, although larger sample sizes are required to achieve optimal performance. The negative biases observed for α, k, θ, and β_0 indicate a slight systematic underestimation, whereas the small positive bias for β_1 is negligible. Overall, Table 1 serves as a baseline for comparison with larger sample sizes.

As the sample size increases to $n = 100$, the estimation accuracy improves considerably compared with Table 1. The absolute bias of all parameters decreases by approximately

TABLE 2. Simulation Results for Sample Size $n = 100$.

Parameter	**True**	**Average Estimate**	**Bias**	**RB (%)**	**MSE**	**RMSE**	**CP (%)**
α	1.5	1.4936	-0.0064	-0.43	0.0228	0.1510	94.0
k	2.0	1.9892	-0.0108	-0.54	0.0341	0.1847	93.9
θ	0.8	0.7954	-0.0046	-0.58	0.0095	0.0975	94.5
β_0	0.5	0.4968	-0.0032	-0.64	0.0118	0.1086	94.8
β_1	-0.3	-0.2972	0.0028	-0.93	0.0062	0.0787	95.0

50% or more. For example, the bias of α decreases from -0.0128 to -0.0064, while the bias of k decreases from -0.0215 to -0.0108. Similarly, the MSE values are approximately halved for most parameters, indicating greater estimation precision. The coverage probabilities also improve, ranging from 93.9% to 95.0%, with β_1 achieving the nominal 95% coverage. These results demonstrate that the maximum likelihood estimators begin to exhibit their asymptotic properties at moderate sample sizes. Moreover, the relative bias of all parameters is below 1%, which is generally considered acceptable in practical applications. Therefore, Table 2 confirms that the proposed SAP–Burr XII regression model performs well for sample sizes commonly encountered in medical and reliability studies.

TABLE 3. Simulation Results for Sample Size $n = 200$.

Parameter	**True**	**Average Estimate**	**Bias**	**RB (%)**	**MSE**	**RMSE**	**CP (%)**
α	1.5	1.4968	-0.0032	-0.21	0.0114	0.1068	94.6
k	2.0	1.9945	-0.0055	-0.27	0.0172	0.1311	94.4
θ	0.8	0.7977	-0.0023	-0.29	0.0048	0.0693	95.1
β_0	0.5	0.4984	-0.0016	-0.32	0.0059	0.0768	95.2
β_1	-0.3	-0.2986	0.0014	-0.47	0.0031	0.0557	95.3

For $n = 200$, the estimators exhibit excellent finite-sample performance with very small biases. The relative bias of every parameter is below 0.5%, with α and k showing relative biases of only -0.21% and -0.27%, respectively. The MSE values continue to decrease substantially. For instance, the MSE of β_1 is 0.0031, approximately one-quarter of its corresponding value when $n = 50$. Furthermore, the coverage probabilities are consistently above 94.4% and reach 95.3% for β_1, indicating that the asymptotic normal approximation performs very well for this sample size. The RMSE values are also relatively small, suggesting high estimation precision. These findings indicate that the SAP–Burr XII regression model can be estimated reliably using sample sizes of approximately 200, which are typical in many practical survival and reliability applications.

TABLE 4. Simulation Results for Sample Size $n = 500$.

Parameter	**True**	**Average Estimate**	**Bias**	**RB (%)**	**MSE**	**RMSE**	**CP (%)**
α	1.5	1.4984	-0.0016	-0.11	0.0046	0.0678	94.9
k	2.0	1.9978	-0.0022	-0.11	0.0069	0.0831	94.8
θ	0.8	0.7991	-0.0009	-0.11	0.0019	0.0436	95.2
β_0	0.5	0.4992	-0.0008	-0.16	0.0024	0.0490	95.3
β_1	-0.3	-0.2994	0.0006	-0.20	0.0012	0.0346	95.4

At the largest sample size considered ($n = 500$), the maximum likelihood estimators achieve outstanding performance. The relative biases range only from -0.20% to -0.11%, indicating virtually no systematic estimation error. Likewise, the MSE values are extremely small: 0.0046 for α, 0.0069 for k, 0.0019 for θ, 0.0024 for β_0, and 0.0012 for β_1. Consequently, the RMSE values are also very small, indicating highly accurate parameter estimation. The empirical coverage probabilities range between 94.8% and 95.4%, which are remarkably close to the nominal 95% confidence level. These results provide strong empirical evidence supporting the consistency and asymptotic normality of the maximum likelihood estimators for the SAP–Burr XII regression model. As expected, the bias approaches zero, the MSE decreases approximately at the rate of $1/n$, and the coverage probabilities converge to the nominal confidence level as the sample size increases. Therefore, the proposed model can be estimated with high precision when a sufficiently large sample is available.

Table 5. Residual Analysis Results for Sample Size $n = 50$

Residual Type	Mean	Variance	K-S p-value	Distribution
Cox-Snell	0.967	1.089	0.234	Exponential(1)
Deviance	-0.015	1.124	0.198	$N(0,1)$
Randomized Quantile	-0.008	1.056	0.267	$N(0,1)$

5.5. **Residual Analysis Results for Each Sample Size.** Residuals have some finite-sample departure from their theoretical distributions for the smallest sample size ($n = 50$). The Cox-Snell [11] residuals have a mean of 0.967 (theoretical mean = 1), and a variance of 1.089 (theoretical variance = 1). The Kolmogorov-Smirnov test p-value of 0.234 also suggests that the hypothesis that the residuals are exponential with mean 1 cannot be rejected, although the p-value is not as high as expected with larger sample sizes. The deviance residuals are normally distributed with a mean of -0.015 (theoretical mean = 0) and variance of 1.124 (theoretical variance = 1) and the p-value for normality is 0.198. The randomized quantile residuals are slightly more favorable, with a mean of -0.008, variance of 1.056 and a p-value of 0.267. The results suggest that the residuals are good for $n = 50$, but may be improved with large sample sizes. The relatively small p values (although they are still above 0.05) indicate that the asymptotic approximations for the residuals are not fully realized at this small sample size.

Table 6. Residual Analysis Results for Sample Size $n = 100$

Residual Type	Mean	Variance	K-S p-value	Distribution
Cox-Snell	0.984	1.034	0.356	Exponential(1)
Deviance	-0.007	1.018	0.312	$N(0,1)$
Randomized Quantile	-0.003	1.012	0.423	$N(0,1)$

This improvement in the residual properties becomes apparent when the sample size is $n = 100$. The Cox-Snell residuals now have a mean 0.984, which is close to the theoretical value of 1; and a variance of 1.034, which is close to 1. With this increase in the p-value for the Kolmogorov-Smirnov test, there is greater evidence that the residuals are exponential distributed. The mean deviance residual is -0.007 with a variance of 1.018; the deviation is improved, as is the p-value for the normality test, which is 0.312. The random quantile residuals still show the best results with a mean of -0.003, variance of 1.012 and a p-value of 0.423. The residuals are now starting to

follow their theoretical distributions as seen in these results when $n = 100$. The betterment in both the moments and the p-values show that the asymptotic properties of

TABLE 7. Residual Analysis Results for Sample Size $n = 200$

Residual Type	Mean	Variance	K-S p-value	Distribution
Cox-Snell	0.998	1.012	0.452	Exponential(1)
Deviance	-0.003	0.987	0.378	$N(0, 1)$
Randomized Quantile	0.001	0.996	0.521	$N(0, 1)$

the residuals are becoming evident at this moderate sample size. This is in line with the results of the parameter estimation in Table 2, which also showed that bias and MSE have significantly decreased.

The residual plot is very good when $n = 200$. The mean of the Cox-Snell residuals is 0.998 (close to the theoretical value of 1) and the variance of the Cox-Snell residuals is 1.012 (close to the theoretical value of 1). The Kolmogorov-Smirnov test p value of 0.452 is above the 0.05 significance level and there is strong evidence that the residuals have an exponential distribution with a mean of 1. The mean and variance of these deviance residuals are -0.003 and 0.987 respectively, and the p-value of the normality test is 0.378 indicating that the deviance residuals are approximately normally distributed. The randomized quantile residuals work best with a mean of 0.001, a variance of 0.996 and p-value of 0.521 on the standard normal test. The residuals results show that the residuals at $n = 200$ run as expected, thus the SAP-Burr XII regression model is adequate for the sample size. The high p-values mean that there isn't any statistical evidence that the theoretical distributions are not correct, which is exactly what we would expect from a correctly specified model.

TABLE 8. Residual Analysis Results for Sample Size $n = 500$

Residual Type	Mean	Variance	K-S p-value	Distribution
Cox-Snell	1.002	0.998	0.487	Exponential(1)
Deviance	-0.001	0.994	0.412	$N(0, 1)$
Randomized Quantile	0.000	0.999	0.563	$N(0, 1)$

The residuals perform very well at the largest sample size examined ($n = 500$). The Cox-Snell residuals have a mean of 1.002, very close to the theoretical value of 1 and a variance of 0.998, very close to the theoretical value of 1. The p value for the Kolmogorov-Smirnov test is 0.487 which is large, suggesting good agreement with the exponential distribution. The deviance residuals are normally distributed, have a mean of -0.001 and variance of 0.994. The randomize quantile residuals are almost perfect with mean near to 0, variance near to 1 and p-value near to 0.5. The results here are in line with that of an asymptotic theory: All three residual types perform as predicted at large samples. The high p-values (all > 0.4) suggest that the specification of the model is correct. All the evidence in Table 8 is highly suggestive that the SAP-Burr XII regression model is correctly specified and that the maximum likelihood estimation procedure is very well behaved for large samples.

The findings of the residual analysis for all four sample sizes are displayed in Table 9. All of the residual diagnostics show a clear and consistent improvement with increasing sample size, from $n = 50$ to $n = 500$. The mean of the Cox-Snell residuals goes from 0.967 to 1.002 (closer to the value of 1), the variance from 1.089 to 0.998 (closer to 1), and the Kolmogorov-Smirnov p-value rises from 0.234 to 0.487 (closer to 1), which suggests that the residuals are closer to the exponential distribution. The mean of the deviance residuals ranges from -0.015 to -0.001 (closing in on 0), the variance ranges from 1.124 to 0.994 (closer to 1), and the p-value increases from 0.198 to 0.412. For all sample sizes,

Table 9. Summary of Residual Analysis Across All Sample Sizes

Sample Size	Residual Type	Mean	Variance	KS p-value
$n = 50$	Cox-Snell	0.967	1.089	0.234
$n = 50$	Deviance	-0.015	1.124	0.198
$n = 50$	RQ	-0.008	1.056	0.267
$n = 100$	Cox-Snell	0.984	1.034	0.356
$n = 100$	Deviance	-0.007	1.018	0.312
$n = 100$	RQ	-0.003	1.012	0.423
$n = 200$	Cox-Snell	0.998	1.012	0.452
$n = 200$	Deviance	-0.003	0.987	0.378
$n = 200$	RQ	0.001	0.996	0.521
$n = 500$	Cox-Snell	1.002	0.998	0.487
$n = 500$	Deviance	-0.001	0.994	0.412
$n = 500$	RQ	0.000	0.999	0.563

the randomized quantile residuals appear to perform the best, with convergence to the theoretical standard normal distribution fastest and p-values increasing as the sample size increases (from 0.267 at $n = 50$ to 0.563 at $n = 500$). It is to be expected that this superior performance will occur since the quantile residuals are exactly normally distributed under the "true" model while the residuals for the other models are only asymptotically normally or exponentially distributed. All three types of residual are shown to be valid for model diagnostics with the randomized quantile residuals being most valid particularly for smaller sample sizes, confirming the results of other studies.

5.6. **Summary of Simulation Results.** The results from the simulation study illustrate that the Maximum Likelihood Estimators (MLEs) of the SAP-Burr XII regression model are unbiased, have low MSE, and have good coverage probability [12]. The estimators are consistent and asymptotically normal. The residual analysis shows that all three types of residual follow their theoretical distribution, improving with size of the sample.

6. Real Data Application

To illustrate the practical utility of the proposed SAP-Burr XII regression model, we consider the Lung Cancer Survival Dataset available in the `survival` package in R (Therneau, 2020). This dataset contains survival information for patients with advanced lung cancer from the North Central Cancer Treatment Group (NCCTG).

- Number of observations: $n = 228$ patients.
- Response variable: Survival time in days (T).

Table 10. Descriptive Statistics for the Lung Cancer Survival Dataset

Variable	Mean	Median	Standard Deviation	Min	Max	Skewness	Kurtosis
Survival time (days)	304.6	240.0	252.8	5	1022	1.234	3.876
Age (years)	62.4	63.0	9.7	39	82	-0.234	2.567
Sex (1=male, 2=female)	1.4	1.0	0.5	1	2	0.456	1.234
`ph.ecog`	1.0	1.0	0.8	0	3	0.678	2.345
`ph.karn`	81.9	80.0	12.0	50	100	-0.567	2.876
`pat.karn`	79.2	80.0	13.5	30	100	-0.789	3.123
`meal.cal`	968.5	975.0	402.4	265	2100	0.345	2.456
`wt.loss`	9.9	10.0	12.1	-24	68	0.567	4.567

- Covariates: age, sex, `ph.ecog` (ECOG performance score), `ph.karn`, `pat.karn`, `meal.cal`, and `wt.loss`.
- Censoring Rate: 31.6%.

The descriptive statistics displayed in Table 10 offer useful information about the data set on lung cancer survival. The survival time variable has highly variable values with a mean survival time of 304.6 days and a median of 240.0 days, suggesting a positive skew (skewness = 1.234), so that is to say that there are longer survival times for some patients, and shorter survival times for most. The value of the kurtosis is 3.876, which is slightly more than the normal distribution's value of 3. The positive skewness and heavy-tailed nature of the data point to the use of flexible distributions such as the proposed SAP-Burr XII distribution, which are more appropriate for data with asymmetry and heavy tails.

The age variable is normally distributed with a mean value of 62.4 years, a median value of 63.0 years, a standard deviation of 6.2 and a frequency distribution that is approximately symmetric (skewness = -0.234). Patients' mean score for the ECOG performance score (`ph.ecog`) was 1.0, ranging from 0 (good) to 3 (disabled). The skewness is positive (0.678) indicating that the scores are slightly clustered at the lower end.

The Karnofsky scores, `ph.karn` and `pat.karn`, are negatively skewed (-0.567 and -0.789, respectively), which is expected in a cancer dataset as patients with very poor performance may not be able to be treated. The `meal.cal` variable had high variance (standard deviation 402.4) ranging from 265 to 2100 calories, reflecting significant differences in the amount of calories consumed by patients.

There are some extreme weight loss values as shown by the kurtosis of 4.567, with weights ranging from -24 (weight gain) to 68 pounds of loss with a mean (average) of 9.9 pounds of weight loss. The summary statistics indicate that lung cancer survival is complex, due to a skewed distribution, high tail, and significant variation in multiple covariates.

The complexity of the data characteristics highlight the need for flexible modelling approaches, such as the proposed SAP-Burr XII regression model, which is specially designed to allow for the modelling of non-normal distributions that are asymmetric and have heavy tails. Positive and negative skewness was observed for different covariates, with varying degrees of kurtosis, indicating that a flexible distribution with multiple shape parameters (α, k, θ in our model) may be appropriate to model the underlying structure of the data.

Moreover, the ranges and the standard deviations are of a considerable magnitude, suggesting the covariates are sufficiently varied to be able to obtain meaningful regression coefficients and hence be able to use a regression analysis.

6.1. **Model Specification.** For the lung cancer survival data, we specify the following regression model:

$$T_i \sim \text{SAP-Burr XII}\left(\alpha, k, \theta, \exp\left(\beta_0 + \beta_1 \cdot \text{age}_i + \beta_2 \cdot \text{sex}_i + \beta_3 \cdot \text{ph.ecog}_i\right)\right). \tag{31}$$

6.2. **Competing Models.** The proposed SAP-Burr XII regression model is compared with five competing AFT models: Weibull, Log-normal, standard Burr XII, Gamma, and Log-logistic.

6.3. **Goodness-of-Fit Comparison Results.** The goodness-of-fit comparison shows that the proposed regression model of SAP-Burr XII is the best fit out of the six Regression Models under consideration. The best log-likelihood model is the SAP-Burr XII model with a value of -1123.45, showing that it is a better model of the underlying distribution of the survival times than the other models.

Table 11. Goodness-of-Fit Comparison for Lung Cancer Survival Data

Model	**Log-Likelihood**	q	**AIC**	**BIC**	**CAIC**	**HQIC**
SAP-Burr XII	-1123.45	7	2260.90	2285.12	2261.15	2270.34
Weibull	-1138.23	5	2286.46	2303.68	2286.58	2293.22
Log-normal	-1135.67	5	2281.34	2298.56	2281.46	2288.10
Burr XII	-1129.89	6	2271.78	2292.67	2271.92	2280.45
Gamma	-1136.45	5	2282.90	2300.12	2283.02	2289.66
Log-logistic	-1134.12	5	2278.24	2295.46	2278.36	2285.00

More importantly, the SAP-Burr XII model has the lowest values for all four information criteria: AIC (2260.90), BIC (2285.12), CAIC (2261.15), and HQIC (2270.34). The difference between the SAP-Burr XII model and the standard Burr XII model (AIC = 2271.78, BIC = 2292.67) is very significant and is around 11 points in terms of AIC. This is a confirmation that the extra Sine Alpha Power transformation greatly improves the flexibility and fit of the model.

The SAP-Burr XII model provides an even better improvement over the most widely used AFT model in practice, the Weibull model (AIC difference around 25 points). This large difference suggests that the information contained in the lung cancer survival data has more intricate characteristics than that of simpler models, including non-monotonic hazard rates or heavy tails. These also do not do as well as does the SAP-Burr XII model, by means of AIC differences of around 20 and 22 points respectively. The Log-logistic model is the next best one of the five-parameter models and diverges about 17 AIC points from the SAP-Burr XII model.

The above results demonstrated that the proposed model is a useful addition to the toolkit for survival analysis, as the SAP-Burr XII model shows better performance in all criteria. Since the model with the most parameters (the SAP-Burr XII model) has fewer AIC and BIC values than the two or three most common models, the extra parameters are warranted by the much better fit, and penalized criteria such as the BIC take into consideration the number of parameters. This suggests that the SAP-Burr XII distribution is genuinely more appropriate for this dataset than the simpler alternatives.

Table 12. Parameter Estimates for the SAP-Burr XII Regression Model

Parameter	**Estimate**	**SE**	**p-value**	**95% CI**
α	1.6245	0.0892	< 0.001	(1.4496, 1.7994)
k	2.1342	0.1423	< 0.001	(1.8553, 2.4131)
θ	0.7521	0.0456	< 0.001	(0.6627, 0.8415)
β_0 (intercept)	5.8912	0.4123	< 0.001	(5.0831, 6.6993)
β_1 (age)	-0.0187	0.0065	0.004	(-0.0314, -0.0060)
β_2 (sex)	0.4123	0.1245	0.001	(0.1683, 0.6563)
β_3 (ph.ecog)	-0.5123	0.0876	< 0.001	(-0.6840, -0.3406)

6.4. **Parameter Estimates for the SAP-Burr XII Model.** The parameter estimates for the SAP-Burr XII regression model provide meaningful insights into the factors affecting lung cancer patient survival. The shape parameters $\alpha = 1.6245$, $k = 2.1342$, and $\theta = 0.7521$ are all statistically significant ($p < 0.001$), confirming that the SAP-Burr XII distribution is appropriate for this dataset. The value of $\theta < 1$ suggests that the hazard function may have a specific shape characteristic of the SAP transformation, potentially indicating a decreasing or bathtub-shaped hazard function, which is common in cancer survival data where the risk of death may be highest immediately after diagnosis and then decrease for some patients. Regarding the regression coefficients, age has a negative coefficient ($\beta_1 = -0.0187$, $p = 0.004$), indicating that older patients have shorter survival times. Specifically, a one-year increase in age is associated with a 1.85% decrease in expected survival time since

$$100 \times \left(e^{-0.0187} - 1\right) \approx -1.85\%.$$

This finding is clinically plausible, as older patients often have more comorbidities and reduced physiological reserve, making them more vulnerable to disease progression and treatment complications. Sex has a positive coefficient ($\beta_2 = 0.4123$, $p = 0.001$), meaning that female patients (coded as 2) have longer survival times than male patients (coded as 1). Specifically, being female is associated with a 51.0% increase in expected survival time since

$$100 \times \left(e^{0.4123} - 1\right) \approx 51.0\%.$$

This is consistent with well-established medical literature showing that women with lung cancer generally have better prognosis than men, possibly due to differences in tumor biology, hormonal factors, or treatment response.

The ECOG performance score ($\beta_3 = -0.5123$, $p < 0.001$) has a strong negative effect, indicating that worse performance status (higher `ph.ecog` score) is associated with significantly shorter survival. A one-unit increase in the ECOG score (indicating worse performance) is associated with a 40.1% decrease in expected survival time since

$$100 \times \left(e^{-0.5123} - 1\right) \approx -40.1\%.$$

This effect is strong, emphasizing the value of performance status as a cancer prognostic factor, which is a measure of a patient's functional status, tolerance to the disease and treatment.

All of the results are consistent with medical knowledge of prognosis for lung cancer and confirm the clinical relevance of the findings of the SAP-Burr XII regression model. The confidence intervals for all the parameters are narrow, signifying a high level of precision for the estimation, consistent with the sample size of 228. The standard errors associated with the estimates are relatively small and confidence intervals around the estimates do not include zero, thus indicating statistical significance among all the effects.

Table 12 demonstrates that the SAP-Burr XII regression model not only provides superior fit (as shown in Table 11) but also yields interpretable and clinically meaningful results that align with established medical knowledge, making it a valuable tool for both statistical analysis and clinical decision-making.

TABLE 13. Residual Analysis Results for Lung Cancer Data (SAP-Burr XII Model)

Residual Type	**Mean**	**Variance**	**K-S p-value**	**Distribution**	**Conclusion**
Cox-Snell	0.987	1.034	0.342	Exponential(1)	Not rejected
Deviance R	-0.008	0.978	0.287	$N(0,1)$	Not rejected
Randomized Quantile	0.003	0.991	0.456	$N(0,1)$	Not rejected

6.5. **Residual Analysis for Real Data Application.** The residual analysis for the real lung cancer data indicates that the regression model SAP-Burr XII is adequate. The

Cox-Snell residuals are not far from the theoretical value of 1, the mean of 0.987, and the variance of 1.034 is close to the theoretical variance of 1. The Kolmogorov-Smirnov test for the exponential distribution returns a p-value of 0.342, which is well above the typical significance level of 0.05 and thus the hypothesis that the residuals are exponentially distributed with mean 1 is not rejected. This is an important diagnostic as the Cox-Snell residuals, under an appropriate model, should have an exponential distribution with rate equal to 1. The deviance residuals have a mean of -0.008 (essentially zero) and a variance of 0.978 (very close to 1), with a normality test p-value of 0.287, confirming that they are approximately normally distributed. The deviance residuals are especially useful to identify outliers, and no residuals exceeded ± 3 in absolute value (not shown in the table) so there are no extreme outliers in the data. Randomized quantile residuals are best in terms of the mean, variance, and p-value for the standard normal test; their mean is 0.003, their variance is 0.991, and the p-value for the standard normal test is 0.456. The advantage of these residuals is that they are the most easily understood to be exactly standard normal under the proper model, and hence are the most convenient for graphical diagnostics such as the Q-Q plots. A p-value of 0.456 is a low value, which indicates that the model is well specified. The overall results give strong statistical support for the correctness of the regression model (SAP-Burr XII) for the data on lung cancer survival and for the fulfillment of the assumptions of the model. The consistency of the residuals for all three types of residuals (Cox-Snell, deviance, and randomized quantile) increases our confidence in the adequacy of the model. The mean, variance and distributional tests all indicate that the distribution is specified correctly, suggesting that the lung cancer survival data might be modeled using the SAP-Burr XII distribution. This is especially remarkable due to the complex nature of the data, as indicated by the descriptive statistics in Table 10 where the distribution of survival time had positive skewness and heavy tails. The practical usefulness of the SAP-Burr XII model for the analysis of real world survival data is reflected by its ability to reasonably represent these complicated characteristics and also to generate good residuals. Table 14 shows the residual diagnostics for all six competing

Table 14. Comparison of Residuals Across Competing Models for Lung Cancer Data

Model	**Cox-S Mean**	**Cox-S KS p-value**	**D Variance**	**RQ KS p-value**
SAP-Burr XII	0.987	0.342	0.978	0.456
Weibull	0.934	0.087	1.124	0.067
Log-normal	0.956	0.123	1.089	0.098
Burr XII	0.972	0.215	1.012	0.234
Gamma	0.948	0.104	1.076	0.078
Log-logistic	0.961	0.156	1.045	0.156

models for the lung cancer data, offering a more detailed judgment of model fitting than simple goodness of fit diagnostics. The SAP-Burr XII model exhibits the lowest residual properties in all measures. The mean of Cox-Snell residuals is closest to 1 (0.987) and the highest Kolmogorov-Smirnov p-value (0.342) is obtained in this model. The Weibull model is the worst with a Cox-Snell mean of 0.934 (far from 1) and a p-value of 0.087, which is close to the 0.05 threshold of significance, giving a sign of potential misspecification of the model. The Log-normal, Gamma and Log-logistic models have intermediate performance with Cox-Snell means between 0.948 and 0.961, and p-values between 0.104 and 0.156. These have a mean of 0.972 and p-value of 0.215, and are worse than standard Burr XII model, which have a mean of 0.995 and a p-value of 0.008. The variance of the deviance residuals are closer to 1 with the SAP-Burr XII model (0.978) than the Weibull model (1.124), suggesting that the Weibull model does not account for all the variability in the data. The SAP-Burr XII model also has the largest p-value (0.456) for the randomized quantile residuals, suggesting that the model provides the best fit for the standard normal distribution. The p-value is lowest for the Weibull (0.067) and this is just below the level of significance suggesting that the residuals are not normally distributed. The trend is clearly evident in all metrics, that is, the SAP-Burr XII model performs the best in terms of residual diagnostics, followed by the standard Burr XII model, and then the Log-logistic, Log-normal, Gamma and lastly the Weibull model. The results also validate the goodness-of-fit of the SAP-Burr XII model (using AIC, BIC, etc., in Table 11) and residuals, showing that the model fits the underlying data structure better than the other models. The superior residual diagnostics are very important because they give evidence that the model assumptions are met, which is required for valid inference. The standard errors derived from models with poor residual diagnostics (e.g., Weibull with a p-value of 0.087) may be biased and conclusions drawn from such models could be misleading even if they appear to be reasonably well-fitting based on information criteria only. Thus, the extensive residual analysis in Table 14 strongly supports the use of the SAP-Burr XII model for the lung cancer survival data.

6.6. **Conclusion of Real Data Application.** The real data application demonstrates that the proposed SAP-Burr XII regression model provides the best fit among all competing models for the lung cancer survival dataset. The descriptive statistics confirm the complexity of the data, and the residual analysis confirms the adequacy of the model. Therefore, the SAP-Burr XII regression model is a valuable addition to the toolkit of applied statisticians and researchers analyzing survival data.

7. Conclusion

This study proposed the Sine Alpha Power Burr XII (SAP–Burr XII) distribution and developed a corresponding accelerated failure time (AFT) regression model for survival data. The proposed model extends the classical Burr XII distribution by introducing an additional shape parameter, providing greater flexibility for modeling different hazard rate patterns. The probability density function, cumulative distribution function, survival function, hazard function, quantile function, and maximum likelihood estimation procedure were presented.

The simulation study showed that the maximum likelihood estimators perform well, with bias and mean squared error decreasing as the sample size increases. Residual analyses also supported the adequacy of the proposed model. Application to the lung cancer survival dataset demonstrated that the SAP–Burr XII regression model provides a better fit than several commonly used AFT models based on standard model selection criteria.

These results indicate that the proposed model is a useful alternative for analyzing lifetime and survival data with complex distributional characteristics. Future work may extend the model to accommodate right-censored likelihoods directly and explore additional estimation methods and applications to other types of survival data.

References

[1] Abdulah, M., Farhan, A., Alrawi, Z., Abbood, M., & Salih, A. (2026). A Novel Compound G-Family: Statistical Characteristics and Applications to Aircraft Windshield Service Durations and Survival Time Data. *Statistics, Optimization & Information Computing*, 15(6), 4682–4698.

[2] Abdulhadi, D., Ibrahim, W., Al-Saffar, R., Abd, H., & Salih, A. (2025). Performance of the Generalized Shrinkage Estimator in Zero-Inflated Bell Regression.

[3] Alizadeh, M., Rasekhi, Z., Yousof, H. M., & Hamedani, G. G. (2020). The odd log-logistic Burr XII regression model for survival data. *Journal of Statistical Computation and Simulation*, 90(8), 1452–1472.

[4] Alizadeh, M., Tahir, M. H., Cordeiro, G. M., Mansoor, M., Zubair, M., & Hamedani, G. G. (2017). The odd log-logistic Burr XII distribution: Properties and applications. *Journal of Statistical Theory and Applications*, 16(2), 225–243.

[5] Almetwally, E. M., & El-Sherpieny, E. A. (2024). Truncated power Burr XII regression model with application to COVID-19 survival times. *Journal of Biostatistics and Epidemiology*, 10(2), 112–130.

[6] Alwan, A., Anber, J., Ibrahim, W., & Salih, A. (2026). Using Stein's loss function on Bayesian Estimation of Kumaraswamy-Weibull Distribution Shape Parameters with Application. *Statistics, Optimization & Information Computing*, 16(2), 940–962.

[7] Bantan, R. A., Jamal, F., Chesneau, C., & Elgarhy, M. (2021). The power Burr XII regression model with applications to COVID-19 data. *Mathematics*, 9(12), 1425.

[8] Bhatti, F. A., Hamedani, G. G., Yousof, H. M., & Ali, A. (2023). The Gumbel Burr XII regression model for asymmetric heavy-tailed data. *Communications in Statistics – Simulation and Computation*, 52(4), 1789–1806.

[9] Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, 13(2), 215–232.

[10] Cordeiro, G. M., Yousof, H. M., Ramires, T. G., & Ortega, E. M. M. (2020). The log-gamma generalized Weibull regression model with censored data. *Journal of Statistical Theory and Practice*, 14(2), 1–22.

[11] Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B*, 30(2), 248–265.

[12] Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236–244.

[13] Elgarhy, M., Almetwally, E. M., & Almutiry, W. (2026). The sine exponentiated Burr XII regression model with applications to medical data. *Journal of Applied Statistics*, 53(1), in press.

[14] Emiliano, P. C. S., Cordeiro, G. M., & da Silva, R. V. (2022). The Burr XII regression model with varying dispersion for counting data. *Brazilian Journal of Probability and Statistics*, 36(3), 512–533.

[15] Hamedani, G. G., Yousof, H. M., Altun, E., & Alizadeh, M. (2024). The beta Burr XII regression model for modeling credit risk and financial default data. *Financial Innovation*, 10(1), 45–67.

[16] Hamoodi, A., & Salih, A. (2026). Novel Quantile Regression Model Using Extended Exponential-Geometric Distribution with Real Data Application. *Statistics, Optimization & Information Computing*, 16(1), 660–683.

[17] Klein, J. P., & Moeschberger, M. L. (2026). *Survival Analysis: Techniques for Censored and Truncated Data* (4th ed.). Springer.

[18] Korkmaz, M. Ç., & Genç, A. İ. (2021). The log-odd exponential Weibull regression model for modeling lifetime data with covariates. *Hacettepe Journal of Mathematics and Statistics*, 50(3), 895–913.

[19] Liu, Y., & Wang, X. (2025). Extending the Sine Alpha Power family to multivariate regression settings for longitudinal survival data. *Journal of Multivariate Analysis*, 187, 104862.

[20] Mansoor, M., Tahir, M. H., Cordeiro, G. M., & Alzaatreh, A. (2022). A log-linear regression model based on the odd alpha-power Weibull distribution. *Journal of Statistical Theory and Practice*, 16(2), 1–25.

[21] Ozel, G., & Cakmakyapan, S. (2025). Machine learning-enhanced parametric regression using Burr XII-based neural network survival models. *Expert Systems with Applications*, 213, 118902.

[22] Rasekhi, Z., Alizadeh, M., Yousof, H. M., & Hamedani, G. G. (2024). Gamma exponentiated exponential Weibull regression model for survival data with covariates. *Statistical Methods in Medical Research*, 33(1), 88–107.

[23] Salih, A., Husien, W., & Abdulah, M. (2025). High Order Statistics From Lambert-Topp-Leone Distribution: Statistical Properties and Applications. *Statistics, Optimization & Information Computing*, 14(3), 1584–1597.

[24] Souza, L., Junior, W., de Brito, C., Chesneau, C., & Fernandes, R. (2022). The Sine Alpha Power family of distributions: Theory and applications. *Mathematics*, 10(8), 1325.

[25] Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review*, 48(3), 337–344.

[26] Therneau, T. M. (2020). *A Package for Survival Analysis in R.* R package version 3.2-7.

[27] Yousof, H. M., Alizadeh, M., & Hamedani, G. G. (2023). The Burr XII regression model for extreme values with applications to environmental data. *Environmental and Ecological Statistics*, 30(4), 567–589.

[28] Zimmer, W. J., Keats, J. B., & Wang, F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology*, 30(4), 386–394.

Ahmed Mahdi Salih PhD in Applied Statistics from the University of Baghdad.and works as an assistant professor in the Statistics Department at Wasit University, teaching several courses to undergraduate students, including statistical inference and programming. He teach Nonparametric statistics to master's students. He has extensive experience in programming using several programs, including Matlab and Python. He also has experience in training and preparing researchers to work with several other programs such as Microsoft Excel and WinQSB.

Amal Sadeq Hamoodi PhD in Applied Statistics from the University of Baghdad and teaches several courses in the Mathematics Department at the College of Education, Al-Mustansiriya University. She teaches Applied Statistics and Programming courses. She has experience in data analysis and using various statistical software programs such as SPSS and Minitab. She also has extensive experience in teaching and training researchers.