## RESEARCH

**Open Access**

CrossMark

# Generalized log-logistic proportional hazard model with applications in survival analysis

Shahedul A. Khan[*] and Saima K. Khosa

*Correspondence: s.khan@usask.ca
Department of Mathematics and
Statistics, University of
Saskatchewan, Wiggins Road,
Saskatoon, SK S7N 5E6, Canada

**Abstract**

Proportional hazard (PH) models can be formulated with or without assuming a probability distribution for survival times. The former assumption leads to parametric models, whereas the latter leads to the semi-parametric Cox model which is by far the most popular in survival analysis. However, a parametric model may lead to more efficient estimates than the Cox model under certain conditions. Only a few parametric models are closed under the PH assumption, the most common of which is the Weibull that accommodates only monotone hazard functions. We propose a generalization of the log-logistic distribution that belongs to the PH family. It has properties similar to those of log-logistic, and approaches the Weibull in the limit. These features enable it to handle both monotone and nonmonotone hazard functions. Application to four data sets and a simulation study revealed that the model could potentially be very useful in adequately describing different types of time-to-event data.

**Keywords:** Cox PH, Log-logistic distribution, Parametric model, Proportional hazard, Semi-parametric model, Time-to-event data, Weibull distribution

**AMS Subject Classification:** Primary 62N01; Secondary 62P10

## Introduction

Proportional hazard (PH) models play a vital role in analyzing time-to-event data. A key assumption in the PH model is that the hazard ratio comparing any two specifications of covariates is constant over time (commonly known as PH assumption). Although the PH assumption may not hold for one or more covariates over the entire study period, it may hold in shorter time intervals. Therefore, violation of the PH assumption may be handled using time-dependent covariates (Kleinbaum and Klein 2012). One of the appealing features of PH models is that the regression coefficients have relative risk interpretation, which is preferred by many clinicians.

The Cox PH model (Cox 1972) is the most popular in survival analysis mainly because of two reasons: (a) no assumption is required about the probability distribution of survival times (i.e., a semi-parametric model), and (b) it usually fits the data well no matter which parametric model is appropriate. In contrast, distributional assumption is required for a fully parametric PH model (Kalbfleisch and Prentice 2002; Lawless 2002). This

also leads to the added requirement of checking the appropriateness of the chosen distribution. Nevertheless, as demonstrated by Efron (1977) and Oakes (1977), parametric models lead to more efficient estimates than Cox's model under certain conditions. More specifically, if the distributional assumption is valid, a parametric model leads to smaller standard errors of the estimates than would be in the absence of a distributional assumption (Collett 2003). Moreover, the use of Cox PH in joint modeling of time-to-event and longitudinal data (Wulfsohn and Tsiatis 1997) usually leads to an underestimation of the standard errors of the parameter estimates (Hsieh et al. 2006; Rizopoulos 2012), and therefore most methods for joint modeling are based on parametric response distributions (Hwang and Pennell 2014). Regarding the choice between a parametric and Cox's PH model, Nardi and Schemper (2003) suggested to use a richer parametric model or simply the Cox's model in case of an unsatisfactory fit of the chosen probability distribution.

The most commonly used parametric time-to-event models are the Weibull, log-logistic and log-normal distributions. The log-logistic and log-normal distributions belong to the accelerated failure time (AFT) family, and are useful in modeling nonmonotone hazard rates (Lawless 2002). Note that the log-logistic also accommodates decreasing hazard functions. Only a few parametric models are closed under PH assumption, the most common of which is the Weibull that accommodates only monotone hazard functions. In fact, Weibull is the only distribution that is closed under both AFT and PH families (Kalbfleisch and Prentice 2002). Mudholkar et al. (1996) proposed a generalization of the Weibull distribution which permits parametric PH regression modeling. It is a three-parameter distribution and is capable of modeling both monotone and non-monotone hazard functions. One difficulty with this model is that it is nonregular (the support depends on some parameters) in the case of increasing hazard functions, and therefore the standard maximum likelihood asymptotics do not hold. In this paper, we propose a simple extension of the log-logistic model which is closed under the PH relationship. The proposed generalized log-logistic model is a three-parameter distribution, and has characteristics similar to those of the log-logistic model. Moreover, it approaches the Weibull in the limit. These features enable it to satisfactorily handle both monotone (increasing and decreasing) and nonmonotone (unimodal) hazard functions. In Section 1, we introduce the generalized log-logistic model and discuss estimation and testing of the parameters using the maximum likelihood method. The proposed method is then illustrated with applications to four data sets, one of which involves joint modeling of time-to-event and longitudinal data (Section Examples). In Section Simulations, a simulation study is presented to evaluate the performance of generalized log-logistic in comparison with other commonly used PH models to describe different types of time-to-event data. We conclude in Section Conclusion by summarizing our findings.

## The generalized log-logistic model

The generalized log-logistic distribution for a nonnegative random variable $T$ can be conveniently specified in terms of the hazard function as follows:

$$h(t; \boldsymbol{\alpha}) = \frac{\kappa \rho (\rho t)^{\kappa-1}}{1 + (\gamma t)^{\kappa}}, t > 0,$$ (1)

where $\rho > 0$, $\kappa > 0$ and $\gamma > 0$ are parameters and $\boldsymbol{\alpha} = (\kappa, \gamma, \rho)'$. If $\gamma$ depends on $\rho$ via $\gamma = \rho$ and $\gamma = \rho \eta^{-1/\kappa}$ with $\eta > 0$, then (1) reduces to the hazard function of the log-logistic (Lawless 2002) and Burr XII (Wang et al. 2008) distributions, respectively. Taking $\gamma$ not dependent on $\rho$, it is easy to verify that (1) is closed under PH relationship (see below). The hazard function is monotone decreasing when $\kappa \leq 1$, and unimodal when $\kappa > 1$ (i.e., $h(t; \boldsymbol{\alpha}) = 0$ at $t = 0$, increases to a maximum at $t = [(\kappa - 1)/\gamma^{\kappa}]^{1/\kappa}$, and then approaches zero monotonically as $t \to \infty$). Note that (1) approaches the Weibull hazard function as $\gamma^{\kappa} \to 0$. This particular feature of the generalized log-logistic model enables it to handle monotone increasing hazard satisfactorily via $\kappa > 1$ and $\gamma$ small (close to zero).

The survivor function, probability density function and cumulative hazard function of the generalized log-logistic distribution are, respectively,

$$S(t; \boldsymbol{\alpha}) = [1 + (\gamma t)^{\kappa}]^{-\frac{\rho^{\kappa}}{\gamma^{\kappa}}}, \tag{2}$$

$$f(t; \boldsymbol{\alpha}) = \frac{\kappa \rho (\rho t)^{\kappa-1}}{[1 + (\gamma t)^{\kappa}]^{\frac{\rho^{\kappa}}{\gamma^{\kappa}}+1}}, \tag{3}$$

$$H(t; \boldsymbol{\alpha}) = \frac{\rho^{\kappa}}{\gamma^{\kappa}} \log[1 + (\gamma t)^{\kappa}]. \tag{4}$$

The median of the distribution is $\dfrac{\left(2^{\frac{\gamma^{\kappa}}{\rho^{\kappa}}} - 1\right)^{\frac{1}{\kappa}}}{\gamma}$, and the $r^{th}$ moment is

$$E(T^r) = \frac{\rho^{\kappa}}{\gamma^{\kappa+r}} \frac{\Gamma\left(\frac{\rho^{\kappa}}{\gamma^{\kappa}} - \frac{r}{\kappa}\right) \Gamma\left(\frac{r}{\kappa} + 1\right)}{\Gamma\left(\frac{\rho^{\kappa}}{\gamma^{\kappa}} + 1\right)} \text{ provided } \frac{\kappa \rho^{\kappa}}{\gamma^{\kappa}} > r.$$

In particular, the mean is $E(T) = \frac{\rho^{\kappa}}{\gamma^{\kappa}} \frac{\Gamma\left(\frac{\rho^{\kappa}}{\gamma^{\kappa}} - \frac{1}{\kappa}\right) \Gamma\left(\frac{1}{\kappa} + 1\right)}{\Gamma\left(\frac{\rho^{\kappa}}{\gamma^{\kappa}} + 1\right)}$ provided $\frac{\kappa \rho^{\kappa}}{\gamma^{\kappa}} > 1$.

For the family of PH models with covariates $\mathbf{z} = (z_1, z_2, \ldots, z_p)'$, the hazard function for $T$ can be expressed as

$$h(t; \mathbf{z}) = h_0(t; \boldsymbol{\alpha}) \, e^{\mathbf{z}' \boldsymbol{\beta}}, \tag{5}$$

where $h_0(t; \boldsymbol{\alpha})$ is the baseline hazard function (i.e., the hazard function when $\mathbf{z} = \mathbf{0}$) characterized by the vector of parameters $\boldsymbol{\alpha}$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is the vector of regression coefficients. A fully parametric PH model can be formulated by specifying $h_0(t; \boldsymbol{\alpha})$ parametrically. If $h_0(t; \boldsymbol{\alpha})$ is specified by the generalized log-logistic hazard function (1), then (5) takes the form

$$h(t; \mathbf{z}) = \frac{\kappa \rho^* (\rho^* t)^{\kappa-1}}{1 + (\gamma t)^{\kappa}}, \tag{6}$$

where $\rho^* = e^{\mathbf{z}' \boldsymbol{\beta}/\kappa}$. Thus the generalized log-logistic is closed under proportionality of hazards. Another widely used parametric PH family is the Weibull, for which $h_0(t; \boldsymbol{\alpha}) = \kappa \rho (\rho t)^{\kappa}$. Note that the Cox PH model is semiparametric, for which the baseline hazard function in (5) is left arbitrary and is denoted by $h_0(t)$.

### Estimation

Suppose that a censored random sample consisting of data $(t_i, \delta_i, \mathbf{z}_i)$, $i = 1, 2, \ldots, n$, is available, where $t_i$ is a lifetime or censoring time according to whether $\delta_i = 1$ or $0$,

respectively, and $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{ip})'$ is the vector of covariates for the $i^{th}$ individual. Letting $m = \sum_{i=1}^{n} \delta_i$, $a_i = \exp(\mathbf{z}_i'\boldsymbol{\beta})$ and $b_i = (\gamma t_i)^\kappa$, the log-likelihood function for the generalized log-logistic PH can be written as

$$\ell(\boldsymbol{\theta}) = m \log \kappa + m\kappa \log \rho + (\kappa - 1) \sum_{i=1}^{n} \delta_i \log t_i - \sum_{i=1}^{n} \delta_i \log(1 + b_i)$$

$$+ \sum_{i=1}^{n} \delta_i \log a_i - \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1 + b_i), \tag{7}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$. The first derivatives of the log-likelihood function are

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa} = \frac{m}{\kappa} + m \log \rho + \sum_{i=1}^{n} \delta_i \log t_i - \frac{1}{\kappa} \sum_{i=1}^{n} \delta_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa}\right) \sum_{i=1}^{n} a_i b_i c_i$$

$$- \left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i \log(1 + b_i), \tag{8}$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} = -\left(\frac{\kappa}{\gamma}\right) \sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right) \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i - \left(\frac{\kappa}{\gamma}\right) \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1 - d_i), \tag{9}$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = \frac{m\kappa}{\rho} - \left(\frac{\kappa}{\rho}\right) \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1 + b_i), \tag{10}$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \delta_i z_{ij} - \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1 + b_i) z_{ij} \text{ for } j = 1, 2, \ldots, p, \tag{11}$$

where $c_i = \log b_i/(1 + b_i)$ and $d_i = b_i/(1 + b_i)$ (see Appendix). To improve the convergence of iterative procedures for maximum likelihood estimation and the accuracy of large-sample methods, we remove range restrictions on parameters through the parameterizations $\boldsymbol{\alpha}^* = (\kappa^*, \gamma^*, \rho^*)'$, where $\kappa^* = \log \kappa$, $\gamma^* = \log \gamma$ and $\rho^* = \log \rho$. The maximum likelihood estimate of $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^{*'}, \boldsymbol{\beta}')'$ can then be obtained by solving the equations $\partial \ell(\boldsymbol{\theta}^*)/\partial \kappa^* = 0$, $\partial \ell(\boldsymbol{\theta}^*)/\partial \gamma^* = 0$, $\partial \ell(\boldsymbol{\theta}^*)/\partial \rho^* = 0$ and $\partial \ell(\boldsymbol{\theta}^*)/\partial \beta_j = 0$ iteratively, where (see Appendix)

$$\frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \kappa^*} = \left[\kappa\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa}\right)\right]_{\boldsymbol{\alpha} = \exp(\boldsymbol{\alpha}^*)}, \quad \frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \gamma^*} = \left[\gamma\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}\right)\right]_{\boldsymbol{\alpha} = \exp(\boldsymbol{\alpha}^*)},$$

$$\frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \rho^*} = \left[\rho\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho}\right)\right]_{\boldsymbol{\alpha} = \exp(\boldsymbol{\alpha}^*)}, \quad \frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \beta_j} = \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j}\right]_{\boldsymbol{\alpha} = \exp(\boldsymbol{\alpha}^*)}.$$

Many software packages have reliable optimization procedures to maximize log-likelihood functions. We wrote our computer code in R (R Core Team 2016), and used the function *nlminb* for optimization (see the Additional file 1).

## Initial values

We may use Weibull, log-logistic and Cox PH fits to generate initial values in solving the equations $\partial \ell(\boldsymbol{\theta}^*)/\partial \kappa^* = 0$, $\partial \ell(\boldsymbol{\theta}^*)/\partial \gamma^* = 0$, $\partial \ell(\boldsymbol{\theta}^*)/\partial \rho^* = 0$ and $\partial \ell(\boldsymbol{\theta}^*)/\partial \beta_j = 0$. Let $\hat{\kappa}_1$ and $\hat{\rho}_1$ be the maximum likelihood estimates of the Weibull shape and scale parameters, respectively, $\hat{\kappa}_2$ and $\hat{\rho}_2$ the maximum likelihood estimates of the log-logistic shape and scale parameters, respectively, and $\hat{\boldsymbol{\beta}}^*$ the estimates of the regression coefficients for the

Cox PH model. Note that maximum likelihood methods for the Weibull, log-logistic and Cox PH models are available in many statistical softwares, including R (R Core Team 2016). We propose to use $\log \hat{\kappa}_1$, $\log |\hat{\kappa}_1 - \hat{\kappa}_2|$, $\log \hat{\rho}_1$ and $\hat{\boldsymbol{\beta}}^*$ as initial values for $\kappa^*$, $\gamma^*$, $\rho^*$ and $\boldsymbol{\beta}$, respectively. If convergence is not achieved with these initial values, we propose to replace $\log \hat{\kappa}_1$ and $\log \hat{\rho}_1$ by $\log \hat{\kappa}_2$ and $\log \hat{\rho}_2$, respectively. In fitting the generalized log-logistic model to many data sets, we have not experienced any difficulty in obtaining convergence with this technique.

### Tests and confidence intervals

Tests and interval estimates for the model parameters are based on the approximate normality of the maximum likelihood estimators. The asymptotic distribution of $\hat{\boldsymbol{\theta}}^*$ is approximately a $(p + 3)$-variate normal distribution with mean $\boldsymbol{\theta}^*$ and covariance matrix $\Sigma = I(\hat{\boldsymbol{\theta}}^*)^{-1}$, where

$$I(\hat{\boldsymbol{\theta}}^*) = - \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^{*2}} & \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^* \partial \gamma^*} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^* \partial \beta_p} \\ \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \gamma^* \partial \kappa^*} & \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \gamma^{*2}} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \gamma^* \partial \beta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \beta_p \partial \kappa^*} & \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \beta_p \partial \gamma^*} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \beta_p^2} \end{bmatrix}_{\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}^*} .$$

is the $(p + 3) \times (p + 3)$ observed information matrix (second derivatives of $\ell(\boldsymbol{\theta}^*)$ are given in Appendix). By the multivariate delta method, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is also approximately normal with mean $\boldsymbol{\theta}$ and covariance matrix $D\Sigma D'$, where $D$ is the $(p + 3) \times (p + 3)$ diagonal matrix $diag(\hat{\boldsymbol{\alpha}}, 1, 1, \ldots, 1)$ and $\hat{\boldsymbol{\alpha}} = \exp(\hat{\boldsymbol{\alpha}}^*)$.

### Generalized log-logistic distribution in joint modeling

Joint models are used to quantify association between an internal time-dependent covariate and time until an event of interest occurs (Wulfsohn and Tsiatis 1997). It involves two separate models: a model that takes into account measurement error in the time-dependent covariate to estimate its true values (longitudinal model), and another model that uses these estimated values to quantify the association between this covariate and the time to the occurrence of the event (time-to-event model). The idea behind the joint modeling technique is to couple the time-to-event model with the longitudinal model. The general framework of the maximum likelihood method and large sample theory can be found in Rizopoulos (2012). Maximization of the log-likelihood function for joint modeling is computationally challenging, as it involves evaluating multiple integrals that do not have an analytical solution, except in very special cases. The R package **JM** has been developed by Rizopoulos (2010) to fit joint models using Weibull baseline hazard, piecewise-constant baseline hazard, spline approximation of the baseline hazard and unspecified baseline hazard functions. We have modified the source codes for Weibull to fit joint models using the generalized log-logistic baseline hazard function. The application of the generalized log-logistic distribution in joint modeling is illustrated with an example in Section 1.

### Goodness of fit

The nonparametric estimates are useful for assessing the quality of fit of a particular parametric time-to-event model (Lawless 2002). For a model without covariate, we use the

approach to simultaneously examine plots of parametric and nonparametric estimates of the survival function, superimposed on the same graph. Let $S(t; \hat{\theta})$ and $\hat{S}(t)$ be the estimates of the survivor functions based on the parametric model of interest and the Kaplan-Meier method (Kaplan and Meier 1958), respectively. The estimates $S(t; \hat{\theta})$ as a function of $t$ should be close to $\hat{S}(t)$ if the parametric model is adequate. For a model with covariates, we consider residual diagnostic plots, where the residuals are defined based on the cumulative hazard function $H(t; \theta)$. If $\hat{S}(H(t; \hat{\theta}))$ is the Kaplan-Meier estimate of $H(t; \hat{\theta})$, then a plot of $-\log \hat{S}(H(t; \hat{\theta}))$ versus $H(t; \hat{\theta})$ should be roughly a straight line with unit slope when the model is adequate (Lawless 2002).

We also use the Akaike's information criterion (AIC) (Akaike 1974) to compare the fits of different models. The AIC is defined by

$$AIC = -2\log(\text{maximized likelihood}) + 2(p + k),$$

where $p$ is the number of covariates and $k$ is the number of parameters of the assumed probability distribution ($k = 3$ for the generalized log-logistic model). In general, when comparing two or more models, we prefer the one with the lowest AIC value. A rule of thumb is that if $\Delta_M = AIC_M - AIC_{min} > 2$, then there is considerably less support for Model M compared to the model with minimum AIC (Burnham and Anderson 2002).

## Examples

Three data sets are taken from the literature to demonstrate the ability of the generalized log-logistic distribution in modeling time-to-event data. The application of the generalized log-logistic PH in joint modeling is illustrated using another data set of AIDS patients. We first use the scaled TTT transform of failure times to detect the shape of the hazard function (Mudholkar et al. 1996). The scaled TTT transform is given by $\phi(v/n) = \left[ \sum_{i=1}^{v} T_{(i)} + (n - v) T_{(v)} \right] / \left( \sum_{i=1}^{n} T_{(i)} \right)$, where $T_{(i)}$ represent the order statistics of the sample, and $v = 1, 2, \ldots, n$. The hazard function is increasing, decreasing and unimodal if the plot of $(v/n, \phi(v/n))$ is concave, convex, and concave followed by convex, respectively. For the first three examples (Sections Example 1: Head and neck cancer data-Example 3: Vaginal cancer mortality in rats), we first fit the generalized log-logistic, Weibull and log-logistic models (without covariate) and check the appropriateness of the distributional assumption using diagnostic plots. Then, we analyze the data using regression models, and compare the fits via residual plots. Note that the regression model based on the log-logistic distribution is given by $\log T = \beta_0 + \beta_1 z_1 + \ldots + \beta_p z_p + \tau W$ where $\tau = 1/\kappa$, $\beta_0 = -\log \rho$ and $W$ has the logistic distribution with density $f(w) = e^w/(1 + e^w)^2$, $-\infty < w < \infty$. This model has an accelerated life interpretation (Lawless 2002), whereas the generalized log-logistic and Weibull PH models have relative risk interpretation. In the fourth example (Section Example 4: AIDS data), we consider joint models based on the generalized log-logistic, Weibull and piecewise-constant baseline hazard functions.

### Example 1: Head and neck cancer data
#### *Data description, hazard shape and distributional assumption*
Efron (1988) described a randomized clinical trial to compare radiation therapy alone (arm A) versus radiation plus chemotherapy (arm B) in treating head and neck cancer patients. Survival times (in days) for 51 patients in arm A (9 observations were censored) and 45 in arm B (14 were censored) were reported. The TTT plot in Fig. 1(a) suggests a
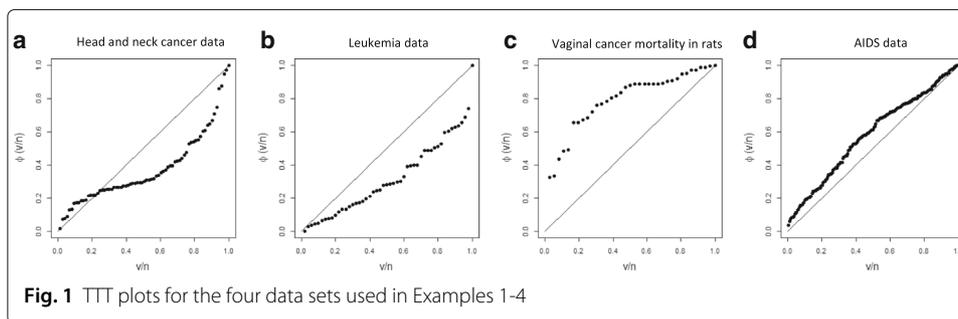
**Fig. 1** TTT plots for the four data sets used in Examples 1-4

unimodal hazard shape of the survival times. Plots of $S(t;\hat{\boldsymbol{\theta}})$ and $\hat{S}(t)$ (Fig. 2(a–c)) indicate more support for the generalized log-logistic distribution in comparison with the Weibull and log-logistic distributions in describing the head and neck cancer data.

### Regression analysis

Letting $z_i = \text{I}(\text{treatment} = \text{radiation therapy})$ that equals 1 if the treatment involves radiation therapy alone and 0 otherwise, we fit the generalized log-logistic PH, Weibull PH and log-logistic AFT models to the head and neck cancer data (numerical results are summarized in Table 1). The standard error of $\hat{\beta}$ for the generalized log-logistic model is smaller that those for the Weibull and log-logistic models, and therefore the generalized log-logistic would be preferred on grounds of efficiency. We also see that the generalized log-logistic has the lowest AIC value, which is supported by the residual plots (Fig. 2(d–f)): residuals lying closely to the unit-slope line for generalized log-logistic indicate its superiority over the Weibull and log-logistic models. In summary, the generalized log-logistic fits the data adequately and is the best among the three models under consideration.
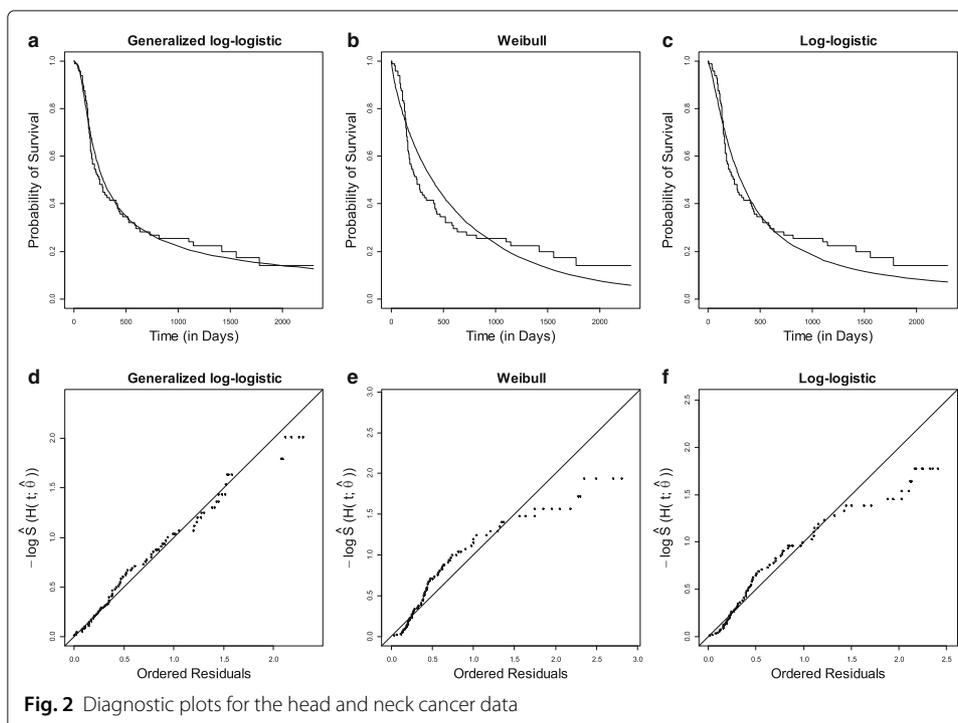


**Fig. 2** Diagnostic plots for the head and neck cancer data

**Table 1** Generalized log-logistic, Weibull and log-logistic fits for the head and neck cancer data

| Parameter | Generalized log-logistic PH (AIC = 1053.39) | | Weibull PH (AIC = 1082.52) | | Log-logistic AFT (AIC = 1067.25) | |
| --- | --- | --- | --- | --- | --- | --- |
| | Estimate | SE | Estimate | SE | Estimate | SE |
| $\beta$ | 0.5459 | 0.2382 | 0.6686 | 0.2415 | −0.5549 | 0.2779 |
| $\log \kappa$ | 0.9790 | 0.1986 | −0.1619 | 0.0921 | 0.2764 | 0.0971 |
| $\log \rho$ | −5.2692 | 0.1844 | −6.8248 | 0.2112 | −6.0492 | 0.2128 |
| $\log \gamma$ | −4.6497 | 0.1755 | | | | |

### Example 2: Autologous and allogeneic bone marrow transplants

#### Data description, hazard shape and distributional assumption

Klein and Moeschberger (2003) described a study involving a sample of 101 patients with advanced acute myelogenous leukemia. Fifty-one of these patients had received an autologous (auto) bone marrow transplant, whereas 50 an allogeneic (allo) transplant. Survival times (in months) for 28 auto transplant and 22 allo transplant patients were censored. Careful inspection of the TTT plot in Fig. 1(b) reveals an indication of the unimodality of the hazard function. A comparison of the diagnostic plots (without covariate) in Fig. 3(a–c) suggests that the assumption of generalized log-logistic is more appropriate than the assumption of Weibull or log-logistic in describing these data.

#### Regression analysis

For regression analysis, we consider the covariate $z_i = $ I(transplant = allo). The fits via the generalized log-logistic PH, Weibull PH and log-logistic AFT are summarized in Table 2. The generalized log-logistic has the lowest AIC value, suggesting it produced the best-fitting model. The residual plots (Fig. 3(d–f)) also support this fact. It is interesting
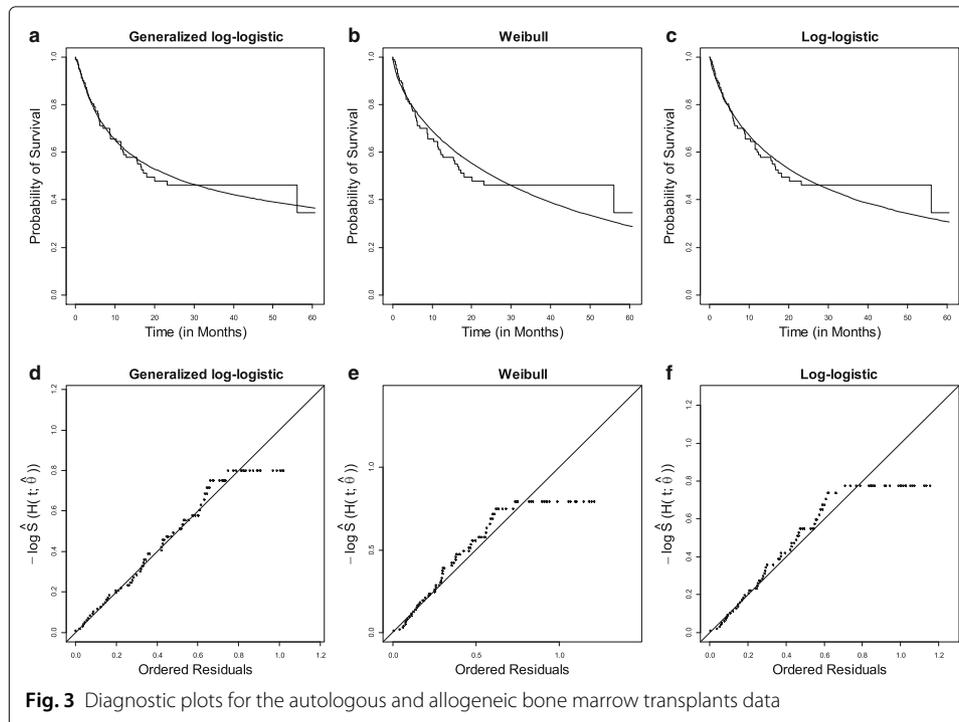


**Fig. 3** Diagnostic plots for the autologous and allogeneic bone marrow transplants data

**Table 2** Generalized log-logistic, Weibull and log-logistic fits for the bone marrow transplants data

| Parameter | Generalized log-logistic PH (AIC = 444.64) | | Weibull PH (AIC = 450.08) | | Log-logistic AFT (AIC = 446.46) | |
|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| $\beta$ | 0.1981 | 0.2854 | 0.2535 | 0.2854 | −0.0808 | 0.4481 |
| $\log \kappa$ | 0.2148 | 0.2376 | −0.3878 | 0.1229 | −0.1694 | 0.1213 |
| $\log \rho$ | −2.4055 | 0.4917 | −3.9683 | 0.3300 | −3.1847 | 0.3474 |
| $\log \gamma$ | −1.3188 | 0.6253 | | | | |

to note here that both the Weibull and log-logistic suggest a decreasing hazard function (estimate of the shape parameter is less than 1), whereas the generalized log-logistic captures the unimodal shape of the hazard function ($\hat{\kappa} = e^{0.2148} = 1.24 > 1$).

### Example 3: Vaginal cancer mortality in rats

#### Data description, hazard shape and distributional assumption

Pike (1966) described a laboratory experiment involving the development of vaginal cancer in rats insulted with the carcinogen DMBA. There were 19 rats in group 1 and 21 in group 2. Seventeen rats in group 1 and 19 in group 2 had developed tumours at the time the data were collected (i.e., two observations in each group were censored). There were reasonable scientific grounds for believing that there might be a threshold value before which no tumour could be detected. For this reason, Lawless (2002) considered the values $t' = t - 100$ to analyze these data. We also consider here the transformed version of the original observations. The TTT plot in Fig. 1(c) suggests an increasing hazard function for $T'$. Figure 4(a–c) shows diagnostic plots for the generalized log-logistic, Weibull and log-logistic fits (without covariate). We see that the generalized log-logistic and Weibull fits are similar, and provide slightly better description of the data compared to the log-logistic model.
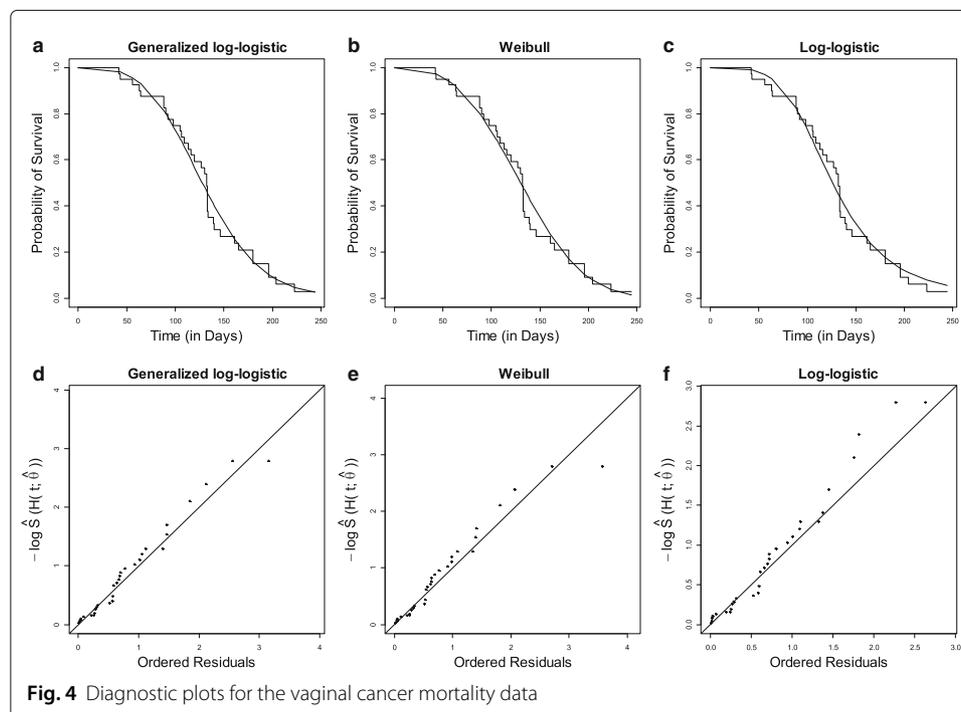


**Fig. 4** Diagnostic plots for the vaginal cancer mortality data

*Regression analysis*

For regression analysis, we consider the covariate $z_i = I(\text{group} = \text{group 1})$. Table 3 gives the estimates of the parameters and associated standard errors from the generalized log-logistic, Weibull and log-logistic fits. As demonstrated by the TTT plot, the Weibull PH suggests an increasing hazard rate ($\hat{\kappa} = e^{1.1308} = 3.098$). Note that a small value of $\hat{\gamma}$ ($\hat{\gamma} = e^{-5.0190} = 0.007$) and $\hat{\kappa} = e^{1.2568} = 3.514 > 1$ for generalized log-logistic also support this fact. Although the AIC values (Table 3) suggest no obvious preference of one model over the other, the residual plots (Fig. 4(d–f)) clearly indicate more support for the generalized log-logistic and Weibull models. This example demonstrates that the generalized log-logistic has the ability to satisfactorily fit data which exhibit increasing hazard rates.

**Example 4: AIDS data**

*Data description and hazard shape*

This example illustrates the use of the generalized log-logistic distribution in joint modeling. Rizopoulos (2012) described a study involving 467 human immunodeficiency virus (HIV) infected patients who had failed or were intolerant to zidovudine therapy (ZT). The main objective was to compare two antiretroviral drugs to prevent the progression of HIV infections: didanosine (ddI) and zalcitabine (ddC). Patients were randomly assigned to receive either ddI or ddC and followed until death or the end of the study, resulting in 188 complete and 279 censored observations. It was also of interest to quantify the association between CD4 cell counts (internal time-dependent covariate) measured at $t = 0$, 2, 6, 12 and 18 months, and time to death. The TTT plot in Fig. 1(d) indicates an increasing hazard shape.

*Regression analysis*

For regression analysis, Rizopoulos (2012) considered joint models of the form

$$h_i(t; \mathbf{z}_i) = h_0(t; \boldsymbol{\alpha}) \exp\{\beta_0 + \beta_1 \text{drug}_i + \beta_2 \text{sex}_i + \beta_3 \text{ZT}_i + \beta_4 \text{CD4}_i(t)\}, \tag{12}$$

$$\text{CD4}_i(t) = b_0 + b_1 t + b_2(t \times \text{drug}_i) + b_{0i} + b_{1i}t + \epsilon_i(t), \tag{13}$$

where (12) is the time-to-event model with $\text{drug}_i = I(\text{drug} = \text{ddI})$, $\text{sex}_i = I(\text{sex} = \text{male})$ and $\text{ZT}_i = I(\text{ZT} = \text{failure})$; and (13) is the longitudinal model with $b_0$, $b_1$ and $b_2$ being the fixed-effects parameters, $b_{0i}$ and $b_{1i}$ the random-effects parameters, and $\epsilon_i(t)$ the random error component. We have reanalyzed the data here using generalized log-logistic, Weibull and piecewise-constant (six knots placed at equally spaced percentiles of the observed event times (Rizopoulos 2012)) baseline hazard functions in (12). Note that

**Table 3** Generalized log-logistic, Weibull and log-logistic fits for the vaginal cancer mortality data

| Parameter | Generalized log-logistic PH (AIC = 391.35) | | Weibull PH (AIC = 389.87) | | Log-logistic AFT (AIC = 391.89) | |
|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| $\beta$ | 0.6254 | 0.3485 | 0.6599 | 0.3474 | −0.1861 | 0.1203 |
| $\log \kappa$ | 1.2568 | 0.2168 | 1.1308 | 0.1300 | 1.5077 | 0.1429 |
| $\log \rho$ | −5.3516 | 0.5889 | −5.0864 | 0.0754 | −4.9301 | 0.0846 |
| $\log \gamma$ | −5.0190 | 0.1154 | | | | |

$h_0(t; \boldsymbol{\alpha}) = \kappa t^{\kappa-1}/[1 + (\gamma t)^\kappa]$ and $\kappa t^{\kappa-1}$ for generalized log-logistic and Weibull, respectively, and so $\beta_0 = \kappa \log \rho$ for both these models. For piecewise-constant baseline hazard, $h_0(t; \boldsymbol{\alpha}) = \sum_{q=1}^{7} \xi_q I(\nu_{q-1} < t \leq \nu_q)$ and $\beta_0 = 0$ in (12), where $0 = \nu_0 < \nu_1 < \ldots < \nu_7$ is the split of the time scale and $\xi_q$ is the value of the hazard in the interval $(\nu_{q-1}, \nu_q]$. The estimates of the parameters and standard errors for the time-to-event process are presented in Table 4. We see that the estimates of the coefficients (i.e., $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$) and their standard errors are broadly similar under the three competing models. The AIC values and residual plots (Fig. 5) also suggest no obvious preference of one model over the other. Although we see no obvious preference of the generalized log-logistic model for this example for which the hazard function is monotone increasing, generalized log-logistic could be useful in joint modeling where the shape of the hazard function is unimodal.

## Simulations

Four covariates in a PH regression framework were considered in all simulations: two continuous covariates ($z_1$ and $z_2$), each generated from the standard normal distribution; and two binary covariates ($z_3$ and $z_4$), each generated from the *Bernoulli*(0.5) distribution. Regression parameter values were chosen to be $\boldsymbol{\beta} = (0.50, -0.50, 0.75, -0.75)'$ corresponding to the covariate vector $\mathbf{z} = (z_1, z_2, z_3, z_4)'$. To evaluate the performance of the generalized log-logistic model, we considered three simulation scenarios based on the shape of the hazard function. For each scenario (see below), lifetime data were generated from the generalized Weibull distribution with probability density function

$$f(t; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \kappa \rho (\rho t)^{\kappa-1} \exp(\mathbf{z}'\boldsymbol{\beta}) \left[1 - \gamma(\rho t)^\kappa\right]^{\frac{\exp(\mathbf{z}'\boldsymbol{\beta}) - 1}{\gamma}}, \tag{14}$$

where $\rho > 0$, $\kappa > 0$ and $-\infty < \gamma < \infty$ are distributional parameters and $\boldsymbol{\alpha} = (\kappa, \gamma, \rho)'$; the support of the distribution is $t > 0$ for $\gamma \leq 0$ and $0 < t < \frac{1}{\rho \gamma^\kappa}$ for $\gamma > 0$. Note that the hazard function of the generalized Weibull distribution is (a) monotone increasing for $\kappa \geq 1$ and $\gamma \geq 1$, (b) monotone decreasing for $0 < \kappa \leq 1$ and $\gamma \leq 1$, and (c) unimodal for $\kappa > 1$ and $\gamma < 0$. The simulation scenarios are then specified as follows.

1. **Scenario 1: Decreasing hazard.** Lifetimes were generated from generalized Weibull with $\kappa = 0.5$, $\gamma = -0.1$ and $\rho = 0.1$, and censoring times were generated from the exponential distribution with rate parameter $\lambda = 0.045$.

**Table 4** AIDS data: estimates and standard errors for the time-to-event process of joint models

| Parameter | Generalized log-logistic (AIC = 8699.61) | | Weibull (AIC = 8699.26) | | Piecewise-constant (AIC = 8711.61) | |
|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| $\beta_0$ | −3.1615 | 0.4411 | −2.9477 | 0.3898 | — | — |
| $\beta_1$ | 0.3690 | 0.1575 | 0.3727 | 0.1576 | 0.3647 | 0.1573 |
| $\beta_2$ | −0.3647 | 0.2583 | −0.3619 | 0.2591 | −0.3364 | 0.2585 |
| $\beta_3$ | 0.3372 | 0.1556 | 0.3455 | 0.1555 | 0.3329 | 0.1555 |
| $\beta_4$ | −0.2824 | 0.0382 | −0.2784 | 0.0378 | −0.2860 | 0.0382 |
| $\log \kappa$ | 0.3838 | 0.7709 | 0.2377 | 0.0732 | | |
| $\log \gamma$ | −2.8874 | 0.1333 | | | | |

**Fig. 5** Residual plots for the AIDS data
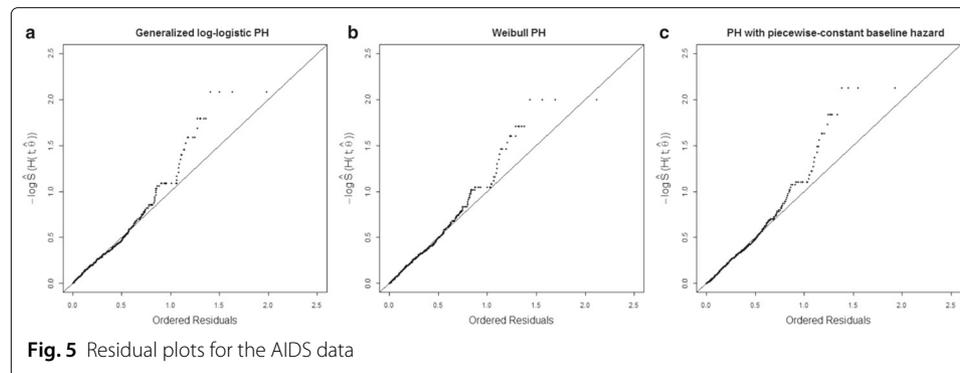
2.  **Scenario 2: Increasing hazard.** Lifetimes were generated from generalized Weibull with $\kappa = 2$, $\gamma = 0.1$ and $\rho = 0.1$, and censoring times were generated from the exponential distribution with rate parameter $\lambda = 0.060$.

3.  **Scenario 3: Unimodal hazard.** Lifetimes were generated from generalized Weibull with $\kappa = 2$, $\gamma = -0.1$ and $\rho = 0.1$, and censoring times were generated from the exponential distribution with rate parameter $\lambda = 0.060$.

Our choice of the parameter values led to, on average, 39.99, 40.69 and 42.99% censored observations for Scenarios 1-3, respectively. Given the covariates and censoring indicator, we then fit the generalized log-logistic, Weibull and Cox PH models to the simulated lifetimes. Note that since the Cox model is robust (usually fits the data well no matter which parametric model is appropriate), we consider this in our simulation study to compare model performance. For each scenario, 500 data sets (each of size $n = 100$) were generated, and the average of each of the estimated model parameters across these data sets was calculated. Absolute bias (AB) and mean square error (MSE) were then computed for model comparison (numerical results are summarized in Table 5).

**Results for scenario 1.** For the continuous covariates ($z_1$ and $z_2$), all three models produced estimates with similar MSE, whereas for the binary covariates ($z_3$ and $z_4$), the generalized log-logistic demonstrated the smallest MSE. In terms of bias, generalized log-logistic and Cox PH were roughly equivalent, and both were superior to Weibull.

**Results for scenario 2.** For the regression coefficients, the generalized log-logistic produced estimates with the smallest bias. We also see that the generalized log-logistic and Weibull produced estimates with similar MSE, and both were superior to the Cox PH model. Note that the generalized log-logistic estimates for $\kappa$ and $\gamma$ were 2.269 and 0.006, respectively (i.e., the estimate of $\gamma^\kappa$ is close to zero), supporting the fact that the hazard function is monotone increasing.

**Results for scenario 3.** In terms of bias, the generalized log-logistic and Cox PH produced comparable estimates of the regression coefficients. However, the generalized log-logistic produced the most accurate estimates in terms of MSE, mostly as a consequence of smaller standard deviations of the estimates. As expected, the Weibull produced the least accurate estimates in terms of both bias and MSE for Scenario 3 (i.e., unimodal hazard).

A simulation study with about 20% censored observations per data set also led to similar conclusions (data not shown). In summary, our simulation study has demonstrated

**Table 5** Model performance and comparison using simulation study ($n = 100$) with about 40% censored observations

| Scenarios | Parameter | True | Generalized log-logistic PH | | | Weibull PH | | | Cox PH | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean | AB | MSE | Mean | AB | MSE | Mean | AB | MSE |
| Scenario 1 | $\beta_1$ | 0.50 | 0.524 | 0.024 | 0.024 | 0.531 | 0.031 | 0.024 | 0.524 | 0.024 | 0.025 |
| (True model: | $\beta_2$ | −0.50 | −0.538 | 0.038 | 0.032 | −0.545 | 0.045 | 0.033 | −0.536 | 0.036 | 0.032 |
| generalized | $\beta_3$ | 0.75 | 0.798 | 0.048 | 0.088 | 0.808 | 0.058 | 0.090 | 0.795 | 0.045 | 0.091 |
| Weibull) | $\beta_4$ | −0.75 | −0.780 | 0.030 | 0.081 | −0.792 | 0.042 | 0.083 | −0.782 | 0.032 | 0.083 |
| | $\rho$ | 0.10 | 0.148 | | | 0.103 | | | | | |
| | $\kappa$ | 0.50 | 0.550 | | | 0.508 | | | | | |
| | $\gamma$ | −0.10 | 0.073 | | | | | | | | |
| Scenario 2 | $\beta_1$ | 0.50 | 0.516 | 0.016 | 0.027 | 0.518 | 0.018 | 0.027 | 0.532 | 0.032 | 0.031 |
| (True model: | $\beta_2$ | −0.50 | −0.522 | 0.022 | 0.035 | −0.523 | 0.023 | 0.035 | −0.533 | 0.033 | 0.039 |
| generalized | $\beta_3$ | 0.75 | 0.785 | 0.035 | 0.099 | 0.788 | 0.038 | 0.099 | 0.813 | 0.063 | 0.112 |
| Weibull) | $\beta_4$ | −0.75 | −0.765 | 0.015 | 0.082 | −0.768 | 0.018 | 0.082 | −0.796 | 0.046 | 0.092 |
| | $\rho$ | 0.10 | 0.107 | | | 0.106 | | | | | |
| | $\kappa$ | 2.00 | 2.269 | | | 2.249 | | | | | |
| | $\gamma$ | 0.10 | 0.006 | | | | | | | | |
| Scenario 3 | $\beta_1$ | 0.50 | 0.519 | 0.019 | 0.025 | 0.530 | 0.030 | 0.026 | 0.516 | 0.016 | 0.026 |
| (True model: | $\beta_2$ | −0.50 | −0.548 | 0.048 | 0.038 | −0.557 | 0.057 | 0.039 | −0.547 | 0.047 | 0.039 |
| generalized | $\beta_3$ | 0.75 | 0.791 | 0.041 | 0.099 | 0.811 | 0.061 | 0.103 | 0.791 | 0.041 | 0.102 |
| Weibull) | $\beta_4$ | −0.75 | −0.790 | 0.040 | 0.089 | −0.811 | 0.061 | 0.093 | −0.792 | 0.042 | 0.095 |
| | $\rho$ | 0.10 | 0.103 | | | 0.098 | | | | | |
| | $\kappa$ | 2.00 | 2.130 | | | 2.016 | | | | | |
| | $\gamma$ | −0.10 | 0.024 | | | | | | | | |

that the generalized log-logistic could potentially be a very useful parametric model to adequately describe different types of time-to-event data.

## Conclusion

In this paper, we proposed a simple extension of the log-logistic distribution to a PH model by appending an additional parameter. As described in Section 1, the proposed model naturally accommodates decreasing and unimodal hazard functions. The log-logistic distribution is known to be useful to describe unimodal hazard functions (Lawless 2002). As demonstrated in Examples 1 and 2, it turns out that the generalized log-logistic may provide better fits in describing unimodal hazard functions compared to the log-logistic distribution. Moreover, our simulation study revealed that the generalized log-logistic could produce more accurate results compared to the Weibull and Cox PH models in describing monotone decreasing and unimodal hazard functions. In summary, the flexibility provided by the generalized log-logistic model could be very useful in adequately describing different types of time-to-event data.

## Appendix: Derivatives of the log-likelihood function

Let $m = \sum_{i=1}^{n} \delta_i$, $a_i = \exp(\mathbf{z}_i' \boldsymbol{\beta})$, $b_i = (\gamma t_i)^\kappa$, $c_i = \log b_i / (1 + b_i)$ and $d_i = b_i / (1 + b_i)$. We have

- $$\log(\gamma t_i) = \frac{\log b_i}{\kappa}, \tag{15}$$

- $$(\gamma t_i)^\kappa \log(\gamma t_i) = \frac{b_i \log b_i}{\kappa}, \tag{16}$$

- $$\frac{\partial b_i}{\partial \kappa} = \frac{\partial}{\partial \kappa} (\gamma t_i)^\kappa = (\gamma t_i)^\kappa \log(\gamma t_i) = \frac{b_i \log b_i}{\kappa}, \tag{17}$$

- $$\frac{\partial \log b_i}{\partial \kappa} = \frac{\log b_i}{\kappa}, \tag{18}$$

- $$\frac{\partial \log(1 + b_i)}{\partial \kappa} = \frac{b_i \log b_i}{\kappa(1 + b_i)} = \frac{b_i c_i}{\kappa}, \tag{19}$$

- $$\frac{\partial b_i \log b_i}{\partial \kappa} = \frac{b_i \log b_i}{\kappa} \log b_i + b_i \frac{\log b_i}{\kappa} = \frac{b_i (\log b_i)(1 + \log b_i)}{\kappa}, \tag{20}$$

- $$\frac{\partial c_i}{\partial \kappa} = \frac{\partial}{\partial \kappa} \frac{\log b_i}{1 + b_i} = \frac{\log b_i}{\kappa(1 + b_i)} \left(1 - \frac{b_i \log b_i}{1 + b_i}\right) = \frac{c_i(1 - b_i c_i)}{\kappa}, \tag{21}$$

- $$\frac{\partial b_i c_i}{\partial \kappa} = \frac{b_i c_i (1 - b_i c_i + \log b_i)}{\kappa} = \frac{b_i c_i (1 + c_i)}{\kappa}, \tag{22}$$

- $$\frac{\partial d_i}{\partial \kappa} = \frac{\partial}{\partial \kappa} \frac{b_i}{1 + b_i} = \frac{b_i \log b_i}{\kappa(1 + b_i)} \left(1 - \frac{b_i}{1 + b_i}\right) = \frac{c_i d_i}{\kappa}, \tag{23}$$

- $$\frac{\partial \log(1 - d_i)}{\partial \kappa} = \frac{\partial}{\partial \kappa} \log(1 + b_i)^{-1} = -\frac{\partial}{\partial \kappa} \log(1 + b_i) = -\frac{b_i c_i}{\kappa}, \tag{24}$$

- $$\frac{\partial b_i}{\partial \gamma} = \frac{\partial}{\partial \gamma}(\gamma t_i)^\kappa = \kappa \gamma^{\kappa-1} t_i^\kappa = \frac{\kappa}{\gamma} b_i, \tag{25}$$

- $$\frac{\partial d_i}{\partial \gamma} = \frac{\partial}{\partial \gamma} \frac{b_i}{1+b_i} = \frac{\kappa}{\gamma} \frac{b_i}{1+b_i}\left(1 - \frac{b_i}{1+b_i}\right) = \frac{\kappa}{\gamma} d_i(1-d_i), \tag{26}$$

- $$\frac{\partial}{\partial \gamma}\log(1-d_i) = -\frac{\partial}{\partial \gamma}\log(1+b_i) = -\frac{\kappa}{\gamma} \frac{b_i}{1+b_i} = -\frac{\kappa}{\gamma} d_i. \tag{27}$$

Using (7) and (15)-(27), we can derive the first and second derivatives of the log-likelihood function as follows.

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa} = \frac{m}{\kappa} + m\log \rho + \sum_{i=1}^{n} \delta_i \log t_i - \frac{1}{\kappa}\sum_{i=1}^{n} \delta_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa}\right) \sum_{i=1}^{n} a_i b_i c_i$$

$$- \left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i \log(1+b_i).$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} = -\left(\frac{\kappa}{\gamma}\right) \sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1-d_i).$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = \frac{m\kappa}{\rho} - \left(\frac{\kappa}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1+b_i).$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \delta_i z_{ij} - \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1+b_i)z_{ij} \text{ for } j=1,2,\ldots,p.$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa^2} = \frac{\partial}{\partial \kappa} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa}$$

$$= -\frac{m}{\kappa^2} + \frac{1}{\kappa^2}\sum_{i=1}^{n} \delta_i b_i c_i - \frac{1}{\kappa^2}\sum_{i=1}^{n} \delta_i b_i c_i(1+c_i) - \left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right)\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i b_i c_i$$

$$+ \left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa^2}\right)\sum_{i=1}^{n} a_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa^2}\right)\sum_{i=1}^{n} a_i b_i c_i(1+c_i)$$

$$- \left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right)\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^\kappa \left\{\log\left(\frac{\rho}{\gamma}\right)\right\}^2 \sum_{i=1}^{n} a_i \log(1+b_i)$$

$$= -\frac{m}{\kappa^2} - \frac{1}{\kappa^2}\sum_{i=1}^{n} \delta_i b_i c_i^2 - \left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa^2}\right)\sum_{i=1}^{n} a_i b_i c_i^2 - \left(\frac{2}{\kappa}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n} a_i b_i c_i$$

$$- \left(\frac{\rho}{\gamma}\right)^\kappa \left\{\log\left(\frac{\rho}{\gamma}\right)\right\}^2 \sum_{i=1}^{n} a_i \log(1+b_i).$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma^2} = \frac{\partial}{\partial \gamma} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}$$

$$= \left(\frac{\kappa}{\gamma^2}\right)\sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right)^2 \sum_{i=1}^{n} \delta_i d_i(1-d_i) + \kappa \rho^\kappa \left(\frac{\kappa+1}{\gamma^{\kappa+2}}\right)\sum_{i=1}^{n} a_i d_i$$

$$- \left(\frac{\kappa}{\gamma}\right)^2 \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i(1-d_i) + \kappa \rho^\kappa \left(\frac{\kappa+1}{\gamma^{\kappa+2}}\right)\sum_{i=1}^{n} a_i \log(1-d_i)$$

$$+ \left(\frac{\kappa}{\gamma}\right)^2 \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i$$

$$= \left(\frac{\kappa}{\gamma^2}\right)\sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right)^2 \sum_{i=1}^{n} \delta_i d_i(1-d_i)$$

$$+ \frac{\kappa(\kappa+1)}{\gamma^2}\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i[d_i + \log(1-d_i)] + \left(\frac{\kappa}{\gamma}\right)^2 \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i^2.$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho^2} = \frac{\partial}{\partial \rho} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = -\frac{m\kappa}{\rho^2} - \frac{\kappa(\kappa-1)}{\rho^2} \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1+b_i).$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_{j'}} = \frac{\partial}{\partial \beta_j} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_{j'}} = -\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1+b_i) z_{ij} z_{ij'} \text{ for } j, j' = 1, 2, \ldots, p.$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa \partial \gamma} = \frac{\partial}{\partial \kappa} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}$$

$$= -\left(\frac{1}{\gamma}\right) \sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{1}{\kappa}\right) \sum_{i=1}^{n} \delta_i c_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa}\right) \sum_{i=1}^{n} a_i c_i d_i$$

$$- \left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i d_i + \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa}\right) \sum_{i=1}^{n} a_i b_i c_i$$

$$- \left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1-d_i) - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i \log(1-d_i)$$

$$= -\left(\frac{1}{\gamma}\right) \sum_{i=1}^{n} \delta_i d_i (1+c_i) - \left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i [d_i + \log(1-d_i) + c_i(d_i - b_i)]$$

$$- \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i [d_i + \log(1-d_i)].$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa \partial \rho} = \frac{\partial}{\partial \kappa} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho}$$

$$= \frac{m}{\rho} - \left(\frac{1}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1+b_i) - \left(\frac{\kappa}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i \log(1+b_i)$$

$$- \left(\frac{\kappa}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \left(\frac{1}{\kappa}\right) \sum_{i=1}^{n} a_i b_i c_i$$

$$= \frac{m}{\rho} - \left(\frac{1}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i [b_i c_i + \log(1+b_i)]$$

$$- \left(\frac{\kappa}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i \log(1+b_i).$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa \partial \beta_j} = \frac{\partial}{\partial \beta_j} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa} = -\left(\frac{1}{\kappa}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i b_i c_i z_{ij} - \left(\frac{\rho}{\gamma}\right)^\kappa \log\left(\frac{\rho}{\gamma}\right) \sum_{i=1}^{n} a_i \log(1+b_i) z_{ij}.$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \rho} = \frac{\partial}{\partial \gamma} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho}$$

$$= \left(\frac{\kappa}{\rho}\right)\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1+b_i) - \left(\frac{\kappa}{\rho}\right)\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i$$

$$= -\left(\frac{\kappa}{\rho}\right)\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1-d_i) - \left(\frac{\kappa}{\rho}\right)\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i$$

$$= -\left(\frac{\kappa}{\rho}\right)\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i [d_i + \log(1-d_i)].$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_j} = \frac{\partial}{\partial \beta_j} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}$$

$$= -\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa} \sum_{i=1}^{n} a_i d_i z_{ij} - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa} \sum_{i=1}^{n} a_i \log(1-d_i) z_{ij}$$

$$= -\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa} \sum_{i=1}^{n} a_i [\, d_i + \log(1-d_i)\,] z_{ij}.$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho \partial \beta_j} = \frac{\partial}{\partial \beta_j} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho}$$

$$= -\left(\frac{\kappa}{\rho}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa} \sum_{i=1}^{n} a_i \log(1+b_i) z_{ij}.$$

The maximum likelihood estimate of $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^{*\prime}, \boldsymbol{\beta}^{\prime})^{\prime}$ is obtained by solving the equations $\partial \ell(\boldsymbol{\theta}^*)/\partial \kappa^* = 0$, $\partial \ell(\boldsymbol{\theta}^*)/\partial \gamma^* = 0$, $\partial \ell(\boldsymbol{\theta}^*)/\partial \rho^* = 0$ and $\partial \ell(\boldsymbol{\theta}^*)/\partial \beta_j = 0$ iteratively. The first and second derivatives of $\ell(\boldsymbol{\theta}^*)$ can be derived by noting that

$$\frac{\partial \ell}{\partial \log u} = u\left(\frac{\partial \ell}{\partial u}\right), \tag{28}$$

$$\frac{\partial^2 \ell}{\partial \log u \, \partial \log v} = u\left(\frac{\partial v}{\partial u}\right)\left(\frac{\partial \ell}{\partial v}\right) + uv\left(\frac{\partial^2 \ell}{\partial u \partial v}\right), \tag{29}$$

$$\frac{\partial^2 \ell}{\partial \log u \, \partial v} = u\left(\frac{\partial^2 \ell}{\partial u \partial v}\right). \tag{30}$$

Using (29)-(30), the first and second derivatives of $\ell(\boldsymbol{\theta}^*)$ can be expressed as

$$\frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \kappa^*} = \left[\kappa\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}}, \quad \frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \gamma^*} = \left[\gamma\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \rho^*} = \left[\rho\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}}, \quad \frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \beta_j} = \left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j}\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^{*2}} = \left[\kappa\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa}\right) + \kappa^2\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa^2}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \gamma^{*2}} = \left[\gamma\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma}\right) + \gamma^2\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma^2}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \rho^{*2}} = \left[\rho\left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho}\right) + \rho^2\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho^2}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \beta_j \partial \beta_{j'}} = \left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_{j'}}\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}}, \quad \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^* \partial \gamma^*} = \left[\kappa\gamma\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa \partial \gamma}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^* \partial \rho^*} = \left[\kappa\rho\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa \partial \rho}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}}, \quad \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \kappa^* \partial \beta_j} = \left[\kappa\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \kappa \partial \beta_j}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \gamma^* \partial \rho^*} = \left[\gamma\rho\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \rho}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}}, \quad \frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \gamma^* \partial \beta_j} = \left[\gamma\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_j}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \rho^* \partial \beta_j} = \left[\rho\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho \partial \beta_j}\right)\right]_{\boldsymbol{\alpha}=e^{\boldsymbol{\alpha}^*}}.$$

## Additional file

## Authors' contributions
SAK, the principal investigator, conceptually developed the proposed distribution with related mathematical results and R code for computation, and drafted the manuscript. SKK contributed in developing mathematical results and writing Sections 1-3. Both authors read and approved the final manuscript.

## Competing interests
The authors declare that they have no competing interests.

## References
Akaike, H: A new look at the statistical model identification. IEEE Trans. Autom. Control. **19**, 716–723 (1974)

Burnham, KP, Anderson, DR: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York (2002)

Collett, D: Modelling Survival Data in Medical Research. Chapman and Hall/CRC, Florida (2003)

Cox, DR: Regression models and life-tables. J. R. Stat. Soc. Ser. B. **34**, 187–220 (1972)

Efron, B: The efficiency of Cox's likelihood function for censored data. J. Am. Stat. Assoc. **72**, 557–565 (1977)

Efron, B: Logistic regression, survival analysis, and the kaplan-meier curve. J. Am. Stat. Assoc. **83**, 414–425 (1988)

Hsieh, F, Tseng, YK, Wang, JL: Joint modeling of survival and longitudinal data: Likelihood approach revisited. Biometrics. **62**, 1037–1043 (2006)

Hwang, BS, Pennell, ML: Semiparametric bayesian joint modeling of a binary and continuous outcome with applications in toxicological risk assessment. Stat. Med. **33**, 1162–1175 (2014)

Kalbfleisch, JD, Prentice, RL: The Statistical Analysis of Failure Time Data. John Wiley & Sons, New Jersey (2002)

Kaplan, EL, Meier, P: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**, 457–481 (1958)

Klein, JP, Moeschberger, ML: Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York (2003)

Kleinbaum, DG, Klein, M: Survival Analysis: A Self-Learning Text. Springer, New York (2012)

Lawless, JF: Statistical Models and Methods for Lifetime Data. John Wiley & Sons, New Jersey (2002)

Mudholkar, GS, Srivastava, DK, Kollia, GD: A generalization of the weibull distribution with application to the analysis of survival data. J. Am. Stat. Assoc. **91**, 1575–1583 (1996)

Nardi, A, Schemper, M: Comparing Cox and parametric models in clinical studies. Stat. Med. **22**, 3597–3610 (2003)

Oakes, D: The asymptotic information in censored survival data. Biometrika. **64**, 441–448 (1977)

Pike, MC: A method of analysis of certain classes of experiments in carcinogenesis. Biometrics. **22**, 142–161 (1966)

R Core Team: R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, R. R Core Team, Vienna (2016)

Rizopoulos, D: Jm: An r package for the joint modelling of longitudinal and time-to-event data. J. Stat. Softw. **35**, 1–33 (2010)

Rizopoulos, D: Joint Models for Longitudinal and Time-to-Event Data With Applications in R. Chapman and Hall/CRC, Florida (2012)

Wang, Y, Hossain, AM, Zimmer, WJ: Useful properties of the three-parameter Burr XII distribution. In: Ahsanullah, M (ed.) Applied Statistics Research Progress, pp. 11–20. Nova Science Publishers, New York, (2008)

Wulfsohn, MS, Tsiatis, AA: A joint model for survival and longitudinal data measured with errror. Biometrics. **53**, 330–339 (1997)