JOURNAL OF APPLIED STATISTICS
2024, VOL. 51, NO. 9, 1772–1791
https://doi.org/10.1080/02664763.2023.2235093

A hierarchical Bayesian analysis for bivariate Weibull distribution under left-censoring scheme

Danielle Peralta[a], Ricardo Puziol de Oliveira[b] and Jorge Alberto Achcar[a]

[a]Ribeir ao Preto Medical School, University of Sao Paulo (USP), Ribeirao Preto, Brazil; [b]Department of Statistics, State University of Maringá (UEM), Maringá, Brazil

ABSTRACT

This paper presents a novel approach for analyzing bivariate positive data, taking into account a covariate vector and left-censored observations, by introducing a hierarchical Bayesian analysis. The proposed method assumes marginal Weibull distributions and employs either a usual Weibull likelihood or Weibull–Tobit likelihood approaches. A latent variable or frailty is included in the model to capture the possible correlation between the bivariate responses for the same sampling unit. The posterior summaries of interest are obtained through Markov Chain Monte Carlo methods. To demonstrate the effectiveness of the proposed methodology, we apply it to a bivariate data set from stellar astronomy that includes left-censored observations and covariates. Our results indicate that the new bivariate model approach, which incorporates the latent factor to capture the potential dependence between the two responses of interest, produces accurate inference results. We also compare the two models using the different likelihood approaches (Weibull or Weibull–Tobit likelihoods) in the application. Overall, our findings suggest that the proposed hierarchical Bayesian analysis is a promising approach for analyzing bivariate positive data with left-censored observations and covariate information.

ARTICLE HISTORY
Received 11 June 2021
Accepted 27 June 2023

KEYWORDS
Weibull distribution; Tobit model; Bayesian analysis; stellar data; left-censored data

1. Introduction

Multilevel data structures are prevalent in various fields, including epidemiology, public health, education, and sociology. For the analysis of survival data within a multilevel framework, one extensively researched class of models is the Cox proportional hazards models with mixed effects, which incorporate cluster-specific random effects modifying the baseline risk function [51,52,66]. The Cox proportional hazards regression model is commonly employed for survival data analysis [1], where random effects are typically introduced to account for within-cluster homogeneity in the outcomes. Also, Cox regression models with mixed effects are useful for analyzing survival data with repeated measures on individuals, individuals nested within some hierarchy, or other scenarios requiring both fixed and random effects. The inclusion of random effects in a Cox proportional hazards model shares similarities with methods for analyzing multilevel data with continuous, binary, or count outcomes [62]. In this way, bivariate Weibull or, as a special case, bivariate

CONTACT Ricardo Puziol de Oliveira rpuziol.oliveira@gmail.com

exponential distributions have been proposed as an alternative for data analysis [40,41,59]. In some cases, the bivariate Weibull distribution is obtained by introducing random factors to capture the dependence between two random variables [24]. Another approach is to use copula functions. Various copula functions have been employed to obtain bivariate Weibull distributions [27], such as the Farlie–Gumbel–Morgenstern (FGM) copula with Weibull marginal distributions, which is referred to as the FGM bivariate Weibull (FGMBW) distribution [4]. The FGMBW distribution, for example, is useful for describing bivariate data with weak correlation between variables in lifetime data.

The main goal of this paper is to introduce a hierarchical Bayesian analysis for bivariate Weibull data. We consider both the usual Weibull likelihood and a Tobit likelihood approach based on marginal Weibull distributions in the presence of left censoring and covariates. The dependence structure between the bivariate data is modeled by the introduction of a frailty or latent variable. The introduction of a random effect that captures the dependence between the responses implies in best fit to the data used in the study (astronomy data) when compared to models assuming independent data. When the proportion of left censoring is very large, the use of usual likelihood techniques in the presence of left censored data (classical or Bayesian approaches) assuming a specified probability distribution for asymmetric positive data (in this study, a Weibull distribution) in general, may not be satisfactory assuming the same probability distribution for the censored and non-censored observations. In this situation, an alternative introduced in the literature would be to assume a Tobit–Weibull model in the data analysis. In the application considered in the study with astronomy data, the proportion of left censorship is not large. Thus the main goal of the study is a comparison of the two methodologies under a hierarchical Bayesian approach. The assumption of a Weibull model and the choice of covariates was considered from a preliminary data analysis that showed an asymmetry of the data. The main reason for the use of the Weibull distribution, usually the most used lifetime distribution in lifetime data applications, is due to the great flexibility of fit for the data. Besides the great flexibility of fit, this model has only two parameters, which implies in great simplicity to get the inferences of interest, especially assuming a left censored scheme.

Many parametric regression models were introduced in the literature to analyze lifetime data in the presence of censored data (see, e.g., [42]). A very popular semi-parametric regression model extensively used in survival data analysis was introduced by Cox [15] assuming proportional hazards (see also [13,16,38,39,43]). In all these models, independent observations are usually assumed, that is, the sample units are not related to each other. However, in some situations, it is possible to have dependent bivariate responses (two or more measurements in the same unit). To capture the correlation between two or more survival times, we could consider the introduction of 'frailties' or latent variables [12,24,54,55,61]. Random effects models are largely used to model heterogeneity as the frailty model introduced by Vaupel [68] used in multivariate survival analysis. Another possibility in the statistical analysis of bivariate lifetime data is to assume existing parametric probability bivariate lifetime distributions as bivariate exponential, bivariate Weibull, bivariate Lindley or bivariate log-normal distributions (see, e.g., [5,7,17,19–21,25,32,33,35,48,49]).

As an example, and motivation for this study, we consider a stellar astronomy bivariate dataset (https://www.iiap. res.in/astrostat/School08/datasets/censor.html) in the presence of left censored observations introduced by Santos *et al.* [60] (see dataset in Appendix 1 at the end of the manuscript). In this example, the authors seek differences in the properties of stars that do and do not host extrasolar planetary systems where a previously identified sample of objects (stars, galaxies, quasars, X-ray sources, etc.) are observed at some new wavelength or for some new property. This data set is related to the birth and death of stars where many questions still exist, despite the scientists now understand over 90% of a star's life [11,14]. Some of the target objects are detected and the value of the new property is measured, while others are not detected. These are assigned as an upper limit to the value of the property based on the uncertainty of the unsuccessful measurement, that is, we have the presence of left-censored data. The probability to find a planet is a steeply rising function of the star's metal content, but it is unclear whether this arises from the metallicity at birth or from later accretion of planetary bodies. The study introduced by Santos *et al.* [60] focuses on two responses associated to the same star: the abundances of the light elements beryllium (Be) and lithium (Li) that are thought to be depleted by internal stellar burning, so that the excess of Be and Li should be present only in the planet accretion scenario of metal enrichment. In this way, we have the presence of left censored bivariate data associated to each star.

Censored data is common in many applications. Usually we have right-censored data especially in medical studies, for example, where we do not know the true survival time for some patients. This could occur when an individual does not experience the event of interest when the study is over; when an individual is lost to follow-up during the study period or when an individual withdraws from the study. The literature presents three major censoring mechanisms: right, left and interval censoring. This study focuses on the censored data on the left, since it occurs when the lower detection limit of an assay is fixed in many fields including stellar astrophysics, biology, chemistry, and environmental sciences.

Several statistical methods have been proposed in the literature to account for left-censoring in cross-sectional (with one measure per subject) or longitudinal (with several measures per subject) studies. Among these methods, we could point out to multiple imputation [22,34,44,56], reverse survival analysis methods [23,31,34,47], quantile regression [26,69] and censored quantile regression [57,58]. Other possibility is to assume the Tobit model with censored outcomes [36,37,45,53,67,71] or by the Buckley–James estimator [8]. Soret *et al.* [63] propose to reverse the Buckley–James least squares algorithm to handle left-censored data enhanced with a Lasso regularization to accommodate high-dimensional predictors.

This paper is organized as follows: Section 2 presents the proposed Weibull model approaches for bivariate data assuming data with left-censoring mechanism and covariates and inference methods for the parameters of the model. Section 3 presents an application of the proposed methodology considering a stellar astronomy data under a hierarchical Bayesian approach. Finally, Section 4 closes this paper with some concluding remarks and directions for future research.

2. Statistical methods

2.1. Weibull likelihood function considering bivariate data in the presence of left-censored data and covariates

Assuming Weibull distributions [70] for the univariate responses, widely known for its simplicity and flexibility in accommodating different forms of hazard function, is the most widely used distribution model for lifetime analysis. The Weibull distribution for a random variable T has probability density function given by

$$f(t) = \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1} \exp\left\{-\left(\frac{t}{\beta}\right)^{\alpha}\right\}, \quad t \geq 0 \tag{1}$$

where α is the shape parameter and β is the scale parameter, both positive. Let us denote the Weibull distribution with density (1) as Wei(α, β). For this distribution, the survival function S(t) = P(T > t) and the hazard function h(t) are given respectively by,

$$S(t) = \exp\left\{-\left(\frac{t}{\beta}\right)^{\alpha}\right\} \quad \text{and} \quad h(t) = \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1} \tag{2}$$

where t $>$ 0 and $\alpha > 0$, $\beta > 0$. The mean of the Weibull distribution with density (1) is given by $E(T) = \beta\Gamma(1 + 1/\alpha)$ where $\Gamma(\cdot)$ denotes the gamma function. In this case, one may have increasing risks (failure rates) if $\alpha > 1$; decreasing if $\alpha < 1$ and constant if $\alpha = 1$, that is, we have great flexibilty of fit for the data.

In the analysis of bivariate data (T_1, T_2) in the presence of a covariate vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ affecting both dependent random variables assuming Weibull distributions Wei(α_1, β_1) and Wei(α_2, β_2), respectively, we consider the use of hierarchical Bayesian methods. In this way, we assume regression models for the scale parameters β_j in the Weibull density (1), given by

$$\beta_{ji} = \exp(\gamma_{j0} + \gamma_{j1}x_{1i} + \gamma_{j2}x_{2i} + \ldots + \gamma_{jp}x_{pi} + w_i) \tag{3}$$

where $\boldsymbol{\gamma_j} = (\gamma_{j0}, \gamma_{j1}, \gamma_{j2}, \ldots, \gamma_{jp})'$ is the regression parameter vector associated to the covariate vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$, $j = 1, 2; i = 1, 2, \ldots, n$ (sample size); w_i is a random factor which captures extra-Weibull variability and dependence structure between both dependent variables (T_1, T_2). The random factors or latent variables (not-observed) $W_i, i = 1, \ldots, n$, are assumed to be independent random variables with a normal N$(0, \sigma^2)$ distribution.

Assuming a left-censored mechanism, the lifetime data is given by $T_j = \max(C_j, Y_j)$ where C_j is a censored time and Y_j is a complete observation, $j = 1, 2$. Define a censorship indicator variable given by $\delta_j = 1$ if T_j is a complete observation $(Y_j > C_j)$ and $\delta_j = 0$ if T_j is a left censored observation $(Y_j \leq C_j)$. In this way, the likelihood function based only in one bivariate observation (t_1, t_2) is given by, $F_1(t_1)^{\delta_1 - 1} f_1(t_1)^{\delta_1} F_2(t_2)^{\delta_2 - 1} f_2(t_2)^{\delta_2}$ where $F_j(t_j) = P(T_j \leq t_j) = 1 - S_j(t_j)$ and $f_j(t_j)$ is the probability density function, $j = 1, 2$.

Thus assuming Weibull distributions Wei(α_1, β_1) and Wei(α_2, β_2) with density (1) for the random variables T_1 and T_2 and the regression models (3) for the scale parameters β_1 and β_2, the likelihood function for the parameters $\alpha_1, \alpha_2, \sigma^2$ and the parameter regression vectors γ_1 and γ_2 in the presence of the fixed covariate vector $\boldsymbol{x}$ and the random factor w_i based on the ith multivariate observation $(t_{1i}, t_{2i}, \delta_{1i}, \delta_{2i})$ is given by

$$\begin{aligned} L(\alpha_1, \alpha_2, \gamma_1, \gamma_2, w_i, \sigma^2) = &\left[1 - \exp\left\{-\left(\frac{t_{1i}}{\beta_{1i}}\right)^{\alpha_1}\right\}\right]^{1-\delta_{1i}} \\ &\times \left[\frac{\alpha_1}{\beta_{1i}^{\alpha_1}} t_{1i}^{\alpha_1 - 1} \exp\left\{-\left(\frac{t_{1i}}{\beta_{1i}}\right)^{\alpha_1}\right\}\right]^{\delta_{1i}} \end{aligned}$$

$$\times\left[1-\exp\left\{-\left(\frac{t_{2i}}{\beta_{2i}}\right)^{\alpha_2}\right\}\right]^{1-\delta_{2i}}$$
$$\times\left[\frac{\alpha_2}{\beta_{2i}^{\alpha_2}}t_{2i}^{\alpha_2-1}\exp\left\{-\left(\frac{t_{2i}}{\beta_{2i}}\right)^{\alpha_1}\right\}\right]^{\delta_{2i}} \tag{4}$$

Inferences for the parameters $\alpha_1, \alpha_2, \gamma_1, \gamma_2$ and $\tau = 1/\sigma^2$ are obtained using a Bayesian hierarchical approach in two stages. For a Bayesian analysis, we could assume uniform or Gamma, G(a, b), prior distributions for the parameters α_1, α_2 with a and b as known hyperparameters where $G(a, b)$ denotes a gamma distribution with mean a/b and variance a/b^2 ; and normal $N(c, d^2)$ prior distributions for the regression parameters $\gamma_{j0}, \gamma_{j1}, \gamma_{j2}, \ldots, \gamma_{jp}, j = 1, 2$ in the first stage of the hierarchical Bayesian approach; in the second stage of the hierarchical Bayesian approach, we assume a gamma prior distribution for the parameter $\tau = 1/\sigma^2$ associated to the normal distribution $N(0, \sigma^2)$ assumed for the random factors $w_i, i = 1, 2, \ldots, n$. Let us denote this model as 'model 1'.

2.2. Tobit models for left-censored data

Another possibility in the data analysis in the presence of left-censored data is to consider a Tobit model [67] that could fit the data by assuming a regression model whose response variable is censored to a prefixed limiting value. The censoring occurs when the response of the regression model is not directly observable, but its independent variables (or covariates) are observed. Tobit models usually assume the normality assumption but could be modeled by other probability distributions (see, e.g., [50]).

If we have a complete observation, that is, $(T > C)$, let us assume a truncated Weibull distribution with probability density function given by

$$f(t \mid T > C) = \frac{f_0(t)}{P(T > C)} \tag{5}$$

where $f_0(t) = \alpha/\beta^\alpha t^{\alpha-1}\exp\{-(t/\beta)^\alpha\}$ and $S_0(t) = P(T > t) = \exp\{-(t/\beta)^\alpha\}$. In this way, let us assume the mixture model, given by the probability density function,

$$f(t) = p\delta C(t) + (1-p)\frac{f_0(t)}{S_0(C)} \tag{6}$$

where $\delta C(t)$ is the Dirac measure at C and p is the associated probability of T to be left-censored for the mixture model and $1-p$ is the probability to be non-censored data. In this case, if $T \le C$, $S(t) = 1$; otherwise, if $T > C$, $S(t) = (1-p)S_0(t)/S_0(C)$ where $S_0(C) = \exp\{-(C/\beta)^\alpha\}$. Observe that for this truncated mixture model the expected value for $T > C$ is given by $E(T) = (1-p)\beta\Gamma(1 + 1/\alpha)/S_0(C)$ where C is fixed (left-censoring). The likelihood function for the parameters p, α and β based on the ith observation is given by

$$L(p, \alpha, \beta/t_i) = p\delta C(t_i) + (1-p)\frac{f_0(t_i)}{S_0(C)} \tag{7}$$

With the censoring information, let us define a binary variable $\delta = 1$ if T is a complete observation $(T > C)$ and $\delta = 0$ if T is a left censored observation $(T \le C)$ with conditional probabilities given by

$$
\begin{aligned}
P(\delta = 0 \mid p, \alpha, \beta, t) &= \frac{p}{p + (1-p)\frac{f_0(t)}{S_0(C)}} \\
P(\delta = 1 \mid p, \alpha, \beta, t) &= \frac{(1-p)\frac{f_0(t)}{S_0(C)}}{p + (1-p)\frac{f_0(t)}{S_0(C)}}
\end{aligned} \tag{8}
$$

In this way, we have a Bernoulli distribution with probability function

$$
P(\delta) = \left[\frac{p}{p + (1-p)\frac{f_0(t)}{S_0(C)}}\right]^{1-\delta} \left[\frac{(1-p)\frac{f_0(t)}{S_0(C)}}{p + (1-p)\frac{f_0(t)}{S_0(C)}}\right]^{\delta} \tag{9}
$$

where $\delta = 1(T > C)$ or $\delta = 0(T \leq C)$. Thus, combining (7) with (9), the likelihood function $L(p, \alpha, \beta)$ based on n observations is given by

$$
L(p, \alpha, \beta / \boldsymbol{t}, \boldsymbol{\delta}) = \prod_{i=1}^{n} p^{(1-\delta_i)} \left[(1-p)\frac{f_0(t_i)}{S_0(C_i)}\right]^{\delta_i} \tag{10}
$$

For our analysis, we assume a truncated Weibull distribution. Moreover, in the analysis of bivariate data in the presence of a covariate vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ affecting both dependent random variables T_1 and T_2, we also assume Weibull distributions Wei(α_1, β_1) and Wei(α_2, β_2), respectively, as considered in Section (2.1). In this way, we assume the same regression models for the scale parameters β_j given by (2.3) and logit models for the parameters p_{ji}, given by,

$$
\operatorname{logit}(p_{ji}) = \log\left(\frac{p_{ji}}{1 - p_{ji}}\right) = \zeta_{j0} + \zeta_{j1}x_{1i} + \zeta_{j2}x_{2i} + \ldots + \zeta_{jp}x_{pi} \tag{11}
$$

for $j = 1, 2; i = 1, 2, \ldots, n$. Observe that we are assuming the same structure for the random factor w_i considered in (3) assuming a normal distribution $N(0, \sigma^2)$ to capture the possible dependence between the two responses. Furthermore, the likelihood function for the parameters $\alpha_1, \alpha_2, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\zeta_1}$ and $\boldsymbol{\zeta_2}$, where $\boldsymbol{\gamma_1} = (\gamma_{10}, \gamma_{11}, \gamma_{12}, \ldots, \gamma_{1p})'$, $\boldsymbol{\gamma_2} = (\gamma_{20}, \gamma_{21}, \gamma_{22}, \ldots, \gamma_{2p})'$, $\boldsymbol{\zeta_1} = (\zeta_{10}, \zeta_{11}, \zeta_{12}, \ldots, \zeta_{1p})'$, $\boldsymbol{\zeta_2} = (\zeta_{20}, \zeta_{21}, \zeta_{22}, \ldots, \zeta_{2p})'$, assuming different left censoring C_i, based on n observations is given by

$$
\begin{aligned}
L(\alpha_1, \alpha_2, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\zeta_1}, \boldsymbol{\zeta_2}) = {} & \prod_{i=1}^{n} p_{1i}^{(1-\delta_{1i})} \left[(1-p_{1i})\frac{f_0(t_{1i})}{S_0(C_{1i})}\right]^{\delta_{1i}} \\
& \times \prod_{i=1}^{n} p_{2i}^{(1-\delta_{2i})} \left[(1-p_{2i})\frac{f_0(t_{2i})}{S_0(C_{2i})}\right]^{\delta_{2i}}
\end{aligned} \tag{12}
$$

where $\delta_{1i} = 1(T_{1i} > C_{1i})$ or $\delta_{1i} = 0(T_{1i} \leq C_{1i})$ and $\delta_{2i} = 1(T_{2i} > C_{2i})$ or $\delta_{2i} = 0(T_{2i} \leq C_{2i})$. For some applications, we could have the same fixed left censoring values in (12), that is, C_1 and C_2 in place of C_{1i} and C_{2i}.

For a hierarchical Bayesian analysis of the model, we assume Gamma(a, b) prior distributions for the parameters α_1 and α_2 and Normal N(c, d^2) prior distributions for the regression parameters $\gamma_{10}, \gamma_{11}, \gamma_{12}, \ldots, \gamma_{1p}; \gamma_{20}, \gamma_{21}, \gamma_{22}, \ldots, \gamma_{2p}; \zeta_{10}, \zeta_{11}, \zeta_{12}, \ldots, \zeta_{1p}$ and $\zeta_{20}, \zeta_{21}, \zeta_{22}, \ldots, \zeta_{2p}$ with a, b, c and d as known hyperparameters in the first stage of the hierarchical Bayesian analysis. In the second stage of the hierarchical Bayesian analysis, we assume the same gamma prior for the parameter $\tau = 1/\sigma^2$ assumed in 'model 1'. Let us denote this model, as 'model 2'. We use MCMC (Markov Chain Monte Carlo) simulation methods to get posterior summaries of interest for the parameters of the models introduced in Sections (2.1) and (2.2) (see, e.g., [10,28–30]).

3. Application to a stellar astronomy dataset

3.1. Classical approach assuming standard polynomial regression models

In this application, we assume a data set related to astronomy introduced by Santos *et al.* [60], who study the light elements lithium (Li), beryllium (Be) and boron (B) as indicators of internal stellar structure and kinematics. Because these elements are destroyed at relatively low temperatures, they give us an idea of how the material inside stars mixes with the hotter interior. Their analysis, together or separately, can provide us with important information about mixing and depletion processes. Many studies of light elements in solar-type stars have been based on the abundance of Li as Li features are easier to measure from high-resolution optical spectra.

On the other hand, using terrestrial telescopes to measure the abundance of Be and B is not a simple task, and only with instruments in space it is possible to obtain the abundances of this element. Despite this difficult, it is important to complement Li studies with analyzes of Be and (if possible) B in solar-type stars. Their abundances can help us probe different regions (depths) within a solar-type star. In special, the authors analyzed the Be abundances for a large sample of field solar-type stars with temperatures in the range 4800–6300 K studying the depletion of Be for dwarfs and sub-giants of different temperatures. In this study, we concentrate in the two responses Li and Be (data set introduced in Appendix 1 at the end of the manuscript).

The data set presents some left-censored observations. We first assume a preliminary data analysis of the data assuming the responses abundance of beryllium (Be) and lithium (Li) as two independent random variables in presence of two covariates Type (Type = 1 indicates planet-hosting stars and Type = 2 is the control sample) and T_{eff} (in degrees Kelvin) is the stellar surface temperature not considering the presence of the censored data (n = 55 uncensored observations for the response Be and n = 36 uncensored observations for the response Li). Figure 1 shows the scatterplots of the responses abundance of beryllium (Be) and lithium (Li) in the logarithm scale versus Type and T_{eff} in the logarithm scale.

From the plots of Figure 1, we observe that the responses abundance of beryllium (Be) are smaller with the control sample (Type = 2) when compared to planet-hosting stars and increases with larger stellar surface temperature T_{eff} in the logarithm scale. We also observe that the responses lithium (Li) show no visible effect of the covariate (Type=2) and increases with larger stellar surface temperature T_{eff} in the logarithm scale. Also, it is observed the presence of possible curvature for the relation of both

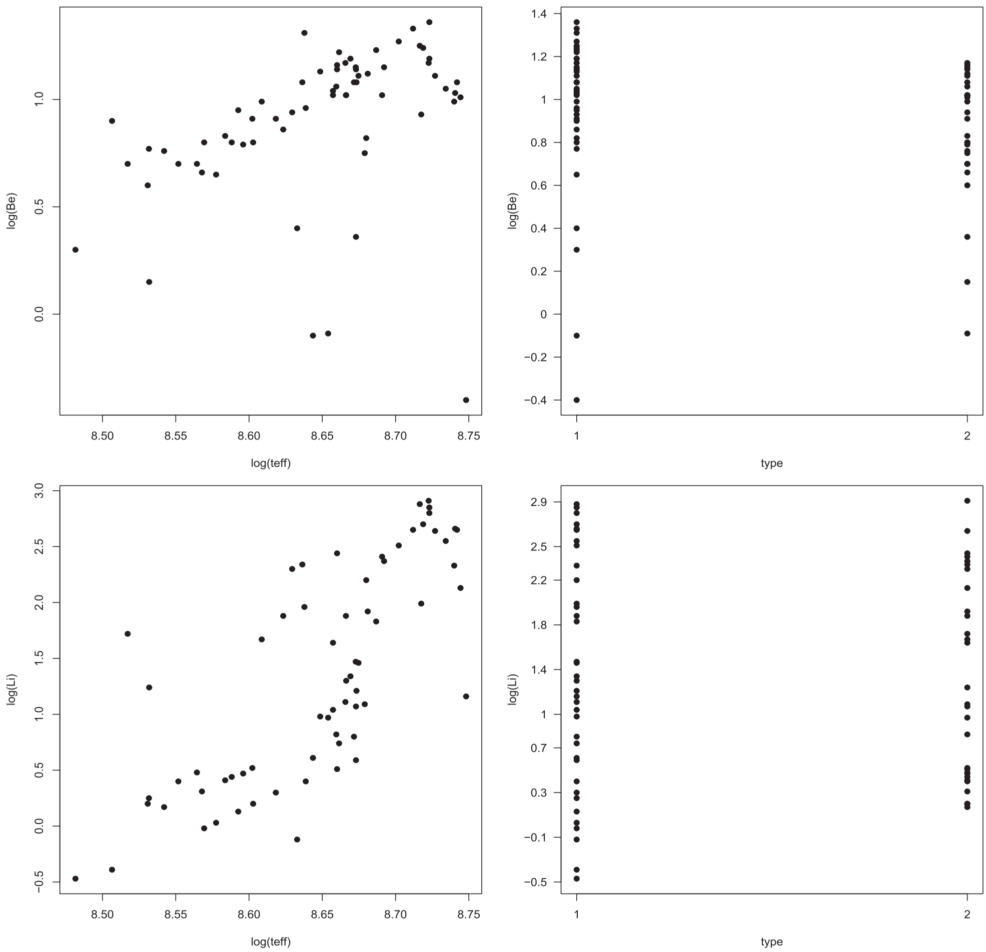

Figure 1. Scatterplots of log(Be) (upper panels) and log(Li) (lower panels) versus Type and log(T_{eff}).

responses Be and Li in logarithm scale versus log(T_{eff}). Assuming linear regression models with standard normal errors with constant variance for the responses abundance of beryllium (Be) and lithium (Li) in logarithm scale in the presence of the covariates type, log(T_{eff}), $[\log(T_{eff})]^2$ and interaction type x log(T_{eff}), Table 2 shows the least square estimators (LSE) for the regression parameters (use of the Minitab software) of the regression models. The needed assumptions for the polynomial regression models (normality and constant variances for the residuals) were verified from residual plots.

Table 1 shows that from a classical linear regression model approach all covariates (type, log(T_{eff}), interaction of type with log(T_{eff}) and quadratic effect of log(T_{eff})) do not have significant effects on the response log(Be) since the p-values > 0.05. Also we observe that assuming a significance level equals to 0.10 the covariates type and interaction type x log(T_{eff}) have significant effects on the response log(Li) since p-value < 0.10 in these cases.

Table 1. LSE for the parameters of the linear regression models.

Source	Coef	SE	T-Value	P-value
Response: log(Be)				
Constant	−673	434	−1.55	0.127
Type	−0.27	7.18	−0.04	0.970
$\log(T_{eff})$	154	100	1.54	0.129
Type x $\log(T_{eff})$	0.019	0.829	0.02	0.982
$\log(T_{eff})^2$	−8.83	5.76	−1.53	0.132
Response: log(Li)				
Constant	750	1694	0.44	0.661
Type	64.1	35.4	1.81	0.079
$\log(T_{eff})$	−193	387	−0.50	0.621
$\log(T_{eff})^2$	12.3	22.1	0.56	0.581
type x $\log(T_{eff})$	−7.35	4.07	−1.81	0.081

3.2. A hierarchical Bayesian analysis assuming the bivariate data in the original scale and left-censoring

In this section, we assume dependent responses abundance of beryllium (Be) and lithium (Li) in the presence of the two covariates Type (Type = 1 indicates planet-hosting stars and Type = 2 is the control sample) and T_{eff} (in degrees Kelvin), the stellar surface temperature, considering all data set presented in Appendix 1, that is, $n = 66$ observations, including the non-censored and the left-censored data in the original scale. We assume Weibull distributions Wei(α_1, β_1) and Wei(α_2, β_2), for the two responses Be and Li with regression models (3) for the scale parameters in presence of the covariates Type and T_{eff} and a random factor W which captures the possible dependence between Be and Li under a hierarchical Bayesian analysis. That is, we assume the regression models given by

$$\begin{aligned}\beta_{1i} &= \exp(\gamma_{10} + \gamma_{11} type_i + \gamma_{12}(\log(T_{eff})_i) + w_i)\\ \beta_{2i} &= \exp(\gamma_{20} + \gamma_{21} type_i + \gamma_{22}(\log(T_{eff})_i)\\ &\quad + \gamma_{23}([\log(T_{eff})_i]^2) + \gamma_{24}(type_i \times \log(T_{eff})_i) + w_i)\end{aligned} \tag{13}$$

where $i = 1, 2, \ldots, 66$; w_i is a random factor which captures extra-Weibull variability and possible dependence between both dependent variables assumed to be independent random variables with a normal N(0, σ^2) distribution. The inclusion of the factors type, $\log(T_{eff})$, $[\log(T_{eff})]^2$ and interaction type x $\log(T_{eff})$ in the regression models for β_1 and β_2 (13) was based from the obtained results in Section 3.1.

For a Bayesian analysis, we assume uniform prior distributions U(0, 10) for the parameters α_1 and α_2; uniform prior distribution U(0, 200) for the parameter $\tau = 1/\sigma^2$; normal prior distributions N(0,1) for the parameters $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \gamma_{23}$ and γ_{24}; and normal prior distributions N(0, 100) for the parameters γ_{10} and γ_{20}. That is, we are assuming approximately non-informative prior distributions for all parameters. We further assume prior independence among the parameters. Inferences for the parameters of the regression models (13) are obtained under a hierarchical Bayesian approach using existing MCMC methods like the Gibbs and the Metropolis–Hastings algorithms.

In the simulation of samples of the joint posterior distribution, $\pi(\boldsymbol{\theta}/data)$ where $\boldsymbol{\theta}$ is the vector of all parameters, we use Gibbs or Metropolis–Hastings algorithms [10,28], where it is needed to sample each parameter from the posterior conditional distributions

Table 2. Posterior summaries for the Weibull regression model (dependent responses).

			95% Cred. Int.	
Parameter	Mean	Std. Dev.	Lower	Upper
α_1	5.0520	0.6863	3.8570	6.4890
α_2	0.9994	0.1469	0.7151	1.2891
γ_{10}	−14.8500	3.9011	−22.0000	−6.9921
γ_{11}	−0.0779	0.0603	−0.1951	0.0528
γ_{12}	1.8470	0.4477	0.9419	2.6710
γ_{20}	−27.0201	7.6620	−41.2400	−11.8702
γ_{21}	−0.2619	0.8979	−2.0400	1.5290
γ_{22}	−1.1260	0.9657	−3.0880	0.6052
γ_{23}	0.5083	0.1312	0.2538	0.7608
γ_{24}	0.0457	0.1105	−0.1761	0.2579
$\tau = 1/\sigma^2$	149.2001	37.2900	66.3201	198.7001

$\pi(\theta_r/\boldsymbol{\theta}(r), data)$, where $\boldsymbol{\theta}(r)$ denotes the vector of all parameters except θ_r and r is associated to each one of the parameters of the model. In this study, we use the OpenBugs software [65] in the simulation of samples of the joint posterior distribution of interest which simplifies the computational work, since this software only requires the definition of the likelihood function for $\boldsymbol{\theta}$ and the prior distribution $\pi(\boldsymbol{\theta})$.

The convergence of the Gibbs sampling algorithm was monitored by usual time series plots for the simulated samples. A burn-in sample of size 111,000 was deleted to eliminate the effects of the initial values in the iterative simulation process and a final Gibbs sample of size 1000 (taken every 100th simulated Gibbs sample) was used to get the posterior summaries of interest. The convergence of the simulation algorithm was verified from trace plots of the simulated Gibbs samples. Table 2 shows the posterior means, posterior standard-deviations and 95% credible intervals for all parameters of the regression models.

Table 2 shows that the stellar surface temperature T_{eff} (in degrees Kelvin) in logarithmic scale, that is, $\log(T_{eff})$ has a significant effect on the response abundance of beryllium (Be) since zero is not included in the 95% credible interval for γ_{12}; the square of the stellar surface temperature T_{eff} (in degrees Kelvin) in logarithmic scale (quadratic effect), that is, $[\log(T_{eff})_i]^2$, has a significant effect on the response abundance of lithium (Li) since zero is not included in the 95% credible interval for γ_{23}. All other covariates do not show significative effects on the responses Be and Li since zero is included in the credible intervals for the corresponding regression parameters. Figure 2 shows the half-normal plots for the residuals of the fitted proposed Weibull regression model from where we can see there is no serious violation of the proposed model.

3.3. A hierarchical Bayesian analysis for the Weibull–Tobbit model assuming the bivariate data in the original scale and left-censoring

As an alternative model, in this section we also assume the dependent responses abundance of beryllium (Be) and lithium (Li) in the original scale with Weibull distributions Wei(α_1, β_1) and Wei(α_2, β_2), respectively, in the presence of the two covariates Type (Type = 1 indicates planet-hosting stars and Type = 2 is the control sample) and T_{eff} (in degrees Kelvin), the stellar surface temperature, considering now the Weibull–Tobit model

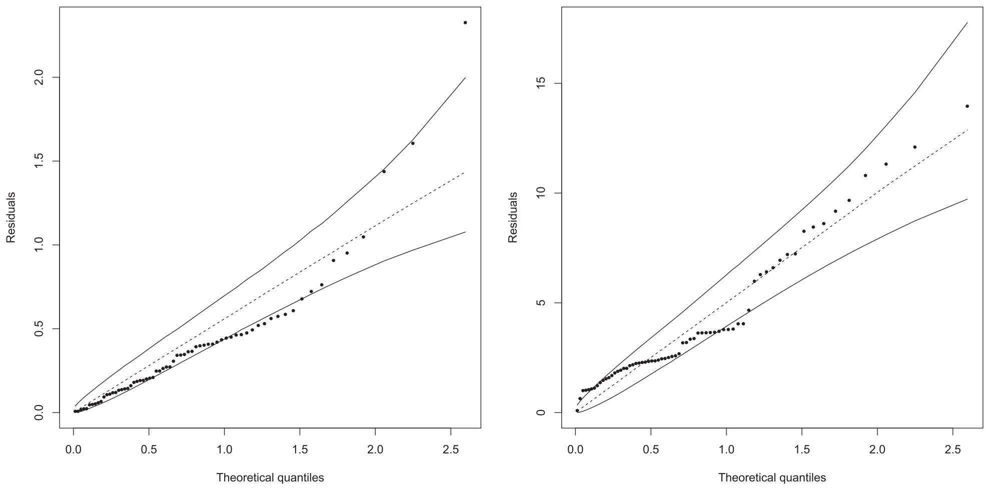

Figure 2. Half-normal plot for residuals of the fitted proposed Weibull regression model for the responses Be (left panel) and Li (right panel).

introduced in Section 3 given by the following regression models:

$$\begin{aligned}\beta_{1i} &= \exp(\gamma_{10} + \gamma_{11} type_i + \gamma_{12}(\log(T_{eff})_i) + w_i)\\ \beta_{2i} &= \exp(\gamma_{20} + \gamma_{21} type_i + \gamma_{22}(\log(T_{eff})_i)\\ &\quad + \gamma_{23}([\log(T_{eff})_i]^2) + \gamma_{24}(type_i \times \log(T_{eff})_i) + w_i)\end{aligned} \tag{14}$$

and

$$\begin{aligned}\text{logit}(p_{1i}) &= \log\left(\frac{p_{1i}}{1-p_{1i}}\right) = \zeta_{10} + \zeta_{11} type_i + \zeta_{12}(\log(T_{eff})_i)\\ \text{logit}(p_{2i}) &= \log\left(\frac{p_{2i}}{1-p_{2i}}\right) = \zeta_{20} + \zeta_{21} type_i + \zeta_{22}(\log(T_{eff})_i\\ &\quad + \zeta_{23}([\log(T_{eff})_i]^2) + \zeta_{24}(type_i \times \log(T_{eff})_i))\end{aligned} \tag{15}$$

where $i = 1, 2, \ldots, n$ (sample size); w_i is a random factor which captures the extra-Weibull variability and dependence between both dependent variables assumed to be independent random variables with a normal N(0, σ^2) distribution.

For a Bayesian analysis, we assume gamma prior distributions G(1,1) for the parameters α_1 and α_2; a uniform prior distribution U(0,100) for the parameter $\tau = 1/\sigma^2$; normal prior distributions N(0,0.01) for the parameters $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$, γ_{23} and γ_{24}; and normal prior distributions N(0,1) for the parameters γ_{10} and γ_{20}; normal prior distributions N(0,0.01) for the parameters $\zeta_{11}, \zeta_{12}, \zeta_{21}, \zeta_{22}$, ζ_{23} and ζ_{24}; and normal prior distributions N(0,1) for the parameters ζ_{10} and ζ_{20}. We further assume prior independence among the parameters. Inferences for the parameters of the regression models above are also obtained under a hierarchical Bayesian approach using existing MCMC methods. It is important to point out that this model (Weibull–Tobit model) has some computational disadvantages when compared to the standard Weibull model in the presence of left-censored data, since the convergence of the MCMC algorithm was only obtained with more informative prior distributions, with small variances.

A burn-in sample of size 111,000 was deleted to eliminate the effects of the initial values in the iterative simulation process and a final Gibbs sample of size 1000 (taking every 50th simulated Gibbs sample) was used to get the posterior summaries of interest. The convergence of the simulation algorithm was verified from trace plots of the simulated Gibbs samples. Table 3 shows the posterior means, posterior standard deviations and 95% credible intervals for all parameters of the regression models. Figure 3 shows the residual plots of the fitted Weibull–Tobit proposed model.

Table 3. Posterior summaries for the Weibull–Tobit model (model 2).

			95% Cred. Int.	
Parameter	Mean	Std. Dev.	Lower	Upper
α_1	8.907	2.028	5.740	13.430
α_2	2.344	0.356	1.669	3.048
γ_{10}	0.099	0.242	−0.321	0.615
γ_{11}	−0.077	0.058	−0.197	0.027
γ_{12}	0.125	0.027	0.071	0.170
γ_{20}	−0.323	0.835	−1.757	1.473
γ_{21}	−0.020	0.086	−0.184	0.152
γ_{22}	−0.008	0.082	−0.176	0.139
γ_{23}	0.036	0.013	0.008	0.059
$\tau = 1/\sigma^2$	14.470	2.972	9.352	20.720
ζ_{10}	−0.862	0.709	−2.223	0.522
ζ_{11}	−0.027	0.099	−0.217	0.169
ζ_{12}	−0.076	0.082	−0.237	0.082
ζ_{20}	0.233	0.982	−1.676	2.229
ζ_{21}	0.019	0.091	−0.157	0.197
ζ_{22}	0.006	0.102	−0.196	0.217
ζ_{23}	−0.001	0.017	−0.041	0.025

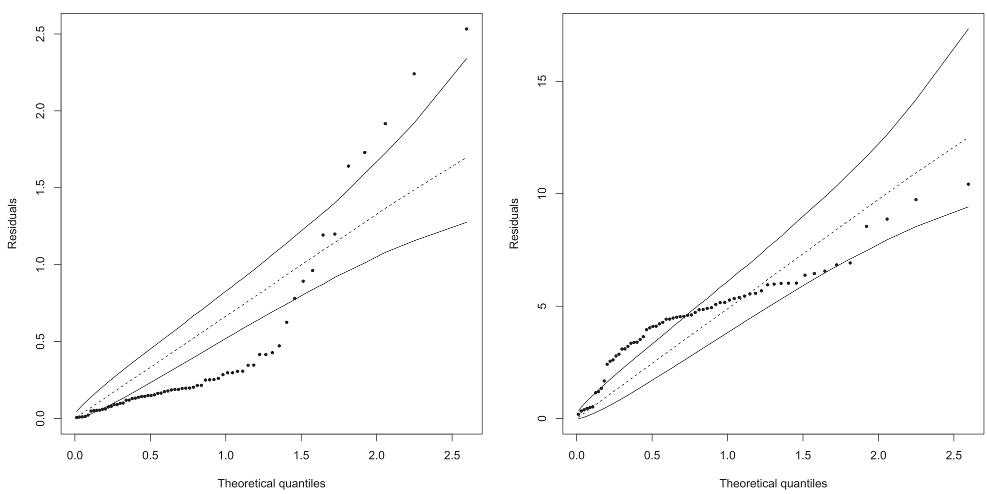

Figure 3. Half-normal plot for residuals of the fitted proposed Weibull–Tobit regression model for the responses Be (left panel) and Li (right panel).

Table 3 shows that the stellar surface temperature T_{eff} (in degrees Kelvin) in logarithmic scale, that is, $\log(T_{eff})$, has a significant effect on the response abundance of beryllium (Be) since zero is not included in the 95% credible interval for γ_{12}; the square of the stellar surface temperature T_{eff} (in degrees Kelvin) in logarithmic scale (quadratic effect), that is, $[\log(T_{eff})_i]^2$, has a significant effect on the response abundance of lithium (Li) since zero is not included in the 95% credible interval for γ_{23}. All other covariates do not show significative effects on the responses Be and Li since zero is included in the credible intervals for the corresponding regression parameters. That is, we have the same conclusions as assuming model 1 (see Table 2).

From the obtained inference results using both assumed models, we have the same inference conclusions, that is, $\log(T_{eff})$, has a significant effect on the response abundance of beryllium (Be) and $[\log(T_{eff})_i]^2$ has a significant effect on the response abundance of lithium (Li). Despite the same conclusions in terms of inferences about the significant covariates affecting the responses of interest, it is observed from the half normal plots of the residuals presented in Figures 2 and 3, a better fit of the bivariate Weibull distribution (half-normal plots in Figure 2) when compared to the bivariate Weibull–Tobit model (half-normal plots in Figure 3), since there are violations of the proposed model in Figure 3. Other possibility is to use more sophisticated discrimination criteria to compare the models assuming independent or dependent responses. A model discrimination extensively used in Bayesian data analysis is the use of the Deviance Information Criterion (DIC) proposed by Spiegelhalter *et al.* [64]. A Deviance Information Criterion (DIC) is a Bayesian measure of model fit that is penalised for complexity similar to the to the Akaike Information Criterion (AIC) extensively used in frequentist data analysis, but many authors have pointed out that the use of DIC could be not satisfactory to discriminate models in the presence of random effects or missing data [6,9]. In this way, to select the best model, we consider the posterior Bayes factor [3], and use the generated Gibbs samples for the parameters of each model to obtain Monte Carlo estimates of the Bayes factor for the different versions of the model. These results are obtained from the OpenBugs software (Monte Carlo estimates for the expected values for the likelihood function assuming each proposed model).

We use the Monte Carlo estimation of the expected value of the likelihood function (or the log-likelihood function) for each model. That would correspond to the values V_i given in an Appendix 2 at the end of the manuscript. Once the values of V_i are obtained for each model, $i = 1,2$, the quantity $B_{12} = V_1/V_2$ may also be obtained and the selection of the best model is performed using the criterion described in Appendix 2. Assuming the Weibull model in the presence of left-censored data we obtain $V_1 = e^{-219.2}$ assuming independent responses and $V_2 = e^{-218.9}$ assuming dependent responses, that is, $B_{12} = 0.740818$ which is an indication that model 2 (dependent responses) is better fitted by the data. Assuming the Tobit–Weibull model, we get $V_3 = e^{-194.2}$, that is, we have an indication that the Tobit–Weibull model with dependent responses is the best fitted model by the data, although we needed to use more informative prior distributions for the parameters of this model, when compared to model 1, to get convergence for the MCMC simulation algorithm using the OpenBUGS software. Other negative point for the assumed Weibull–Tobit model is the large number of parameters assuming logistic regression models for the unknown probabilities $\mathbf{p}_i$ associated to be the left censored data assuming the mixture model. In terms of parcimony, the use of the standard Weibull model in the presence of left-censored data is reasonable in the data analysis (similar inferences results obtained using the two models indicating the same significant covariates affecting the bivariate responses). It is important to point out that the posterior Bayes factors usually indicates the model with more parameters as it is the case of the assumed Weibull–Tobit model.

4. Concluding remarks

The main goal of this paper was to introduce a hierarchical Bayesian analysis of bivariate lifetime data in the presence of left-censored data and covariates based on marginal Weibull distributions. The dependence structure between the two responses was considered by the introduction of a frailty or latent variable. The hierarchical Bayesian analysis was assumed considering the standard Weibull likelihood and the Weibull–Tobit likelihood in the presence of left-censored data and covariates. An illustration of the proposed methodology was considered assuming a stellar astronomy bivariate data set introduced in Appendix 1.

In the application with the astronomy data, we observed that the obtained Bayesian inference results implied in similar results considering both proposed models, in terms of discovering the significative effects of the covariates Type (Type = 1 indicates planet-hosting stars and Type = 2 is the control sample) and T_{eff} (in degrees Kelvin) is the stellar surface temperature on both astronomy responses abundance of beryllium (Be) and lithium (Li) and with similar computational costs to simulate samples for the joint posterior distributions of interest using the free OpenBugs software, although the model 2 (Weibull–Tobit model) presented some convergence problems of the MCMC iterative approach to generate samples for the joint posterior distribution of interest. In this case, the convergence of the simulation algorithm was only obtained assuming more closed prior distributions for the parameters of the model using the OpenBugs software.

In the application considered in this study as an illustration of the proposed methodology, with astronomy data we observed using the standard posterior Bayes factor in the discrimination of the two proposed models (Weibull likelihood in the presence of left-censored data with covariates and Tobit–Weibull likelihood in the presence of left-censored data and covariates) that the models with the introduction of a random effect which captures the dependence structure between the responses led to better fit of the data, in comparison for the use of Weibull models assuming independent responses. From the study carried out, we can conclude that model 1 (bivariate Weibull model obtained by introducing a latent variable in the regression structure for the scale parameter) is a better model when compared to the proposed bivariate Weibull–Tobit model despite the result of the posterior Bayes factor which indicated the bivariate Weibull–Tobit model as the best model in the analysis of the astronomy data set. Considering model 1 (bivariate Weibull model), a sensitivity analysis considering different prior structures was made, leading to good convergence of the MCMC algorithm in all cases and similar posterior summaries. Considering model 2 (bivariate-Weibull distribution), the convergence of the MCMC algorithm was only obtained assuming special classes of informative prior distributions using the OpenBugs software. Other applications and possibly, some simulation studies, should be considered in future studies to better compare the adequability and performance of the two proposed models (models 1 and 2) in each application.

The great advantage of the proposed hierarchical Bayesian methodology in the analysis of bivariate data is the simple form of the likelihood given by product of the likelihood functions and the dependence structure given by a non-observed latent factor or frailty which also could be generalized to other structures. It is interesting to observe that using existing bivariate parametric distributions or parametric distributions derived from copula functions, the likelihood function usually has more computational cost. The likelihood function assuming a continuous model (see, e.g., Lawless, 1982, page 479) is given by

$$L = \prod_{i \in C_1} f(t_{i1}, t_{i2}) \prod_{i \in C_2} \left(-\frac{\partial S(t_{1i}, t_{2i})}{\partial t_{1i}} \right) \prod_{i \in C_3} \left(-\frac{\partial S(t_{1i}, t_{2i})}{\partial t_{2i}} \right) \prod_{i \in C_4} S(t_{i1}, t_{i2}) \tag{16}$$

where $f(t_{i1}, t_{i2})$ is the joint probability function for T_{1i} and T_{2i}; $S(t_{i1}, t_{i2})$ is the joint survival function; $\partial S(t_{1i}, t_{2i})/\partial t_{1i}$ and $\partial S(t_{1i}, t_{2i})/\partial t_{2i}$ are the partial derivatives of $S(t_{i1}, t_{i2})$ with respect to t_{1i} and t_{2i}, respectively.

Other point, especially in applications, in favor of our approach: the use of parametric bivariate probability models derived from copula functions, usually depends on the choice of a particular copula function among hundreds of existing copula functions, since each copula represents different dependence structure for the data set. It is interesting to point out that despite the problems presented assuming the Weibull–Tobit model (lack of convergence with the MCMC algorithms if we do not assume informative prior distributions for the parameters of the model, presence of many parameters, lack of identifiability for the estimation of the parameter p if we do not consider covariates with a logistic structure) in the analysis of the stellar data, the Weibull–Tobit model could give better interpretations of interest to researchers. Usually, mixture models as considered in the Tobit model given by (7), have some advantages in the interpretations, in the same way as obtained with the use of cure fraction models where it is possible to get estimator for susceptible and non-susceptible individuals that can die from some diseases [2,18,46].

In addition, other existing parametric lifetime distributions as exponential, gamma, log-normal or generalizations of the Weibull distribution could be considered to model the univariate distributions for the two responses of the bivariate data in presence of left-censored data. Finally, it is important to point out that the use of existing Bayesian simulation software like the OpenBugs software implies in great simplification to obtain the Bayesian inferences of interest. Another advantage of the Bayesian methodology: it is possible to use expert opinion in the elicitation of prior distributions that can lead to more accurate inference results.

Acknowledgements

The authors thank the referees and editor for the important suggestions that led to the great improvement of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] O. Aalen, O. Borgan, and H. Gjessing, *Survival and Event History Analysis: a Process Point of View*, Springer Science & Business Media, 2008.
[2] J.A. Achcar, E.A. Coelho-Barros, and J. Mazucheli, *Cure fraction models using mixture and non-mixture models*, Tatra Mt. Math. Publ. 51 (2012), pp. 1–9.
[3] M. Aitkin, *Posterior Bayes factors*, J. R. Stat. Soc. B Methodol. 53 (1991), pp. 111–128.
[4] E.M. Almetwally, H.Z. Muhammed, and E.S.A. El-Sherpieny, *Bivariate Weibull distribution: properties and different methods of estimation*, Ann. Data Sci. 7 (2020), pp. 163–193.
[5] B.C. Arnold and D. Strauss, *Bivariate distributions with exponential conditionals*, J. Am. Stat. Assoc. 83 (1988), pp. 522–527.
[6] N. Best, A. Mason, and S. Richardson, *Two-pronged strategy for using dic to compare selection models with non-ignorable missing responses*, Bayesian Anal. 7 (2012), pp. 109–146.
[7] H.W. Block and A. Basu, *A continuous, bivariate exponential extension*, J. Am. Stat. Assoc. 69 (1974), pp. 1031–1037.
[8] J. Buckley and I. James, *Linear regression with censored data*, Biometrika 66 (1979), pp. 429–436.
[9] G. Celeux, F. Forbes, C.P. Robert, and D.M. Titterington, *Deviance information criteria for missing data models*, Bayesian Anal. 1 (2006), pp. 651–673.
[10] S. Chib and E. Greenberg, *Understanding the Metropolis-Hastings algorithm*, Am. Stat. 49 (1995), pp. 327–335.
[11] C. Chiosi, *Fundamentals of stellar evolution theory: Understanding the hrd*, *Stellar astrophysics for the local group: VIII Canary Islands Winter School of Astrophysics*, 1998, p. 1.
[12] D. Clayton and J. Cuzick, *Multivariate generalizations of the proportional hazards model*, J. R. Stat. Soc. A General. 148 (1985), pp. 82–108.
[13] D. Collett, *Modelling Survival Data in Medical Research*, 2nd ed., Chapman and Hall, New York, 2003.
[14] G.W. Collins, *The fundamentals of stellar astrophysics*, WH Freeman and Co., New York, 1989, p. 512 .
[15] D.R. Cox, *Regression models and life tables*, J. R. Stat. Soc. B 34 (1972), pp. 187–220.
[16] D.R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman & Hall, London, 1984.
[17] R.P. de Oliveira, J.A. Achcar, D. Peralta, and J. Mazucheli, *Discrete and continuous bivariate lifetime models in presence of cure rate: a comparative study under bayesian approach*, J. Appl. Stat. 46 (2019), pp. 449–467.
[18] R.P. de Oliveira, A.F. Menezes, J. Mazucheli, and J.A. Achcar, *Mixture and nonmixture cure fraction models assuming discrete lifetimes: application to a pelvic sarcoma dataset*, Biom. J. 61 (2019), pp. 813–826.
[19] R.P. de Oliveira, *Multivariate lifetime models to evaluate long-term survivors in medical studies*, Ph.D. diss., University of São Paulo, 2019.
[20] R.P. de Oliveira, M.V. de Oliveira Peres, J.A. Achcar, and N. Davarzani, *Inference for the trivariate Marshall-Olkin-Weibull distribution in presence of right-censored data.*, Chil. J. Stat. 11 (2020), pp. 95–116.
[21] R.P. de Oliveira, J.A. Achcar, J. Mazucheli, and W. Bertoli, *A new class of bivariate Lindley distributions based on stress and shock models and some of their reliability properties*, Reliab. Eng. Syst. Saf. 211 (2021), pp. 107528.
[22] F. Del Greco M, C. Pattaro, C. Minelli, and J.R. Thompson, *Bayesian analysis of censored response data in family-based genetic association studies*, Biom. J. 58 (2016), pp. 1039–1053.
[23] G.E. Dinse, T.A. Jusko, L.A. Ho, K. Annam, B.I. Graubard, I. Hertz-Picciotto, F.W. Miller, B.W. Gillespie, and C.R. Weinberg, *Accommodating measurements below a limit of detection: a novel application of cox regression*, Am. J. Epidemiol. 179 (2014), pp. 1018–1024.
[24] C.A. dos Santos and J.A. Achcar, *A bayesian analysis for the block and basu bivariate exponential distribution in the presence of covariates and censored data*, J. Appl. Stat. 38 (2011), pp. 2213–2223.
[25] F. Downton, *Bivariate exponential distributions in reliability theory*, J. R. Stat. Soc. B Methodol. 32 (1970), pp. 408–417.

[26] P.H. Eilers, E. Röder, H.F. Savelkoul, and R.G. van Wijk, *Quantile regression for the statistical analysis of immunological data with many non-detects*, BMC. Immunol. 13 (2012), pp. 1–8.
[27] A.Q. Flores, *Testing copula functions as a method to derive bivariate Weibull distributions*, in *American Political Science Association (APSA), Annual Meeting*. Citeseer, 2009, pp. 1–19.
[28] A.E. Gelfand and A.F. Smith, *Sampling-based approaches to calculating marginal densities*, J. Am. Stat. Assoc. 85 (1990), pp. 398–409.
[29] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin, *Bayesian Data Analysis*, CRC press, 2013.
[30] W. Gilks, S. Richardson, and D. Spieglhalter, *Markov chain monte carlo in practice*, 1996, p. 520.
[31] B.W. Gillespie, Q. Chen, H. Reichert, A. Franzblau, E. Hedgeman, J. Lepkowski, P. Adriaens, A. Demond, W. Luksemburg, and D.H. Garabrant, *Estimating population distributions when some data are below a limit of detection by using a reverse kaplan-meier estimator*, Epidemiology (2010), pp. S64–S70.
[32] E.J. Gumbel, *Bivariate exponential distributions*, J. Am. Stat. Assoc. 55 (1960), pp. 698–707.
[33] A.G. Hawkes, *A bivariate exponential distribution with applications to reliability*, J. R. Stat. Soc. B Methodol. 34 (1972), pp. 129–131.
[34] D.R. Helsel, *More than obvious: better methods for interpreting nondetect data*, Environ. Sci. Technol. 39 (2005), pp. 419A–423A.
[35] P. Hougaard, *A class of multivariate failure time distributions*, Biometrika 73 (1986), pp. 671–678.
[36] J.P. Hughes, *Mixed effects models with censored data with application to hiv rna levels*, Biometrics 55 (1999), pp. 625–629.
[37] H. Jacqmin-Gadda, R. Thiébaut, G. Chêne, and D. Commenges, *Analysis of left-censored longitudinal data with application to viral load in hiv infection*, Biostatistics 1 (2000), pp. 355–368.
[38] J.D. Kalbfleisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data*, 2nd ed., Wiley, New York, NY, 2002.
[39] J.P. Klein and M.L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, Springer-Verlag, New York, 1997.
[40] D. Kundu and A.K. Dey, *Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm*, Comput. Stat. Data. Anal. 53 (2009), pp. 956–965.
[41] D. Kundu and A.K. Gupta, *Bayes estimation for the Marshall–Olkin bivariate Weibull distribution*, Comput. Stat. Data. Anal. 57 (2013), pp. 271–281.
[42] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1982.
[43] E.T. Lee and J.W. Wang, *Statistical Methods for Survival Data Analysis*, 3rd ed., Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003.
[44] M. Lee, L. Kong, and L. Weissfeld, *Multiple imputation for left-censored biomarker data based on gibbs sampling method*, Stat. Med. 31 (2012), pp. 1838–1848.
[45] H.S. Lynn, *Maximum likelihood inference for left-censored hiv rna data*, Stat. Med. 20 (2001), pp. 33–45.
[46] R.A. Maller and X. Zhou, *Survival Analysis with Long-Term Survivors*, Wiley, New York, 1996.
[47] I.C. Marschner, R.A. Betensky, V. DeGruttola, S.M. Hammer, and D.R. Kuritzkes, *Clinical trials using hiv-1 rna-based primary endpoints: statistical analysis and potential biases.*, J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. Offic. Publ. Int. Retrovirol. Assoc. 20 (1999), pp. 220–227.
[48] A.W. Marshall and I. Olkin, *A multivariate exponential distribution*, J. Am. Stat. Assoc. 62 (1967), pp. 30–44.
[49] A.W. Marshall and I. Olkin, *A family of bivariate distributions generated by the bivariate bernoulli distribution*, J. Am. Stat. Assoc. 80 (1985), pp. 332–338.
[50] G. Martínez-Flórez, H. Bolfarine, and H.W. Gómez, *The alpha-power Tobit model*, Commun. Stat. Theory Methods. 42 (2013), pp. 633–643.
[51] C. McGilchrist and C. Aisbett, *Regression with frailty in survival analysis*, Biometrics 47 (1991), pp. 461–466.

[52] M. Mills, *Introducing Survival and Event History Analysis*, Sage, 2010.
[53] L. Nie, H. Chu, C. Liu, S.R. Cole, A. Vexler, and E.F. Schisterman, *Linear regression with an independent variable subject to a detection limit*, Epidemiology (Cambridge, Mass.) 21 (2010), pp. S17–S24.
[54] D. Oakes, *Semiparametric inference in a model for association in bivanate survival data*, Biometrika 73 (1986), pp. 353–361.
[55] D. Oakes, *Bivariate survival models induced by frailties*, J. Am. Stat. Assoc. 84 (1989), pp. 487–493.
[56] W.B. Paxton, R.W. Coombs, M.J. McElrath, M.C. Keefer, J. Hughes, F. Sinangil, D. Chernoff, L. Demeter, B. Williams, and L. Corey, et al. *Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-infected subjects with $\geq$ 400 cd4 lymphocytes: implications for applying measurements to individual patients*, J. Infect. Dis. 175 (1997), pp. 247–254.
[57] J.L. Powell, *Least absolute deviations estimation for the censored regression model*, J. Econom. 25 (1984), pp. 303–325.
[58] J.L. Powell, *Censored regression quantiles*, J. Econom. 32 (1986), pp. 143–155.
[59] C.A. Santos and J.A. Achcar, *A Bayesian analysis in the presence of covariates for multivariate survival data: an example of application*, Rev. Colomb. Estad. 34 (2011), pp. 111–131.
[60] N. Santos, G. Israelian, R.G. López, M. Mayor, R. Rebolo, S. Randich, A. Ecuvillon, and C.D. Cerdeña, *Are beryllium abundances anomalous in stars with giant planets?*, Astron. Astrophys. 427 (2004), pp. 1085–1096.
[61] J.H. Shih, *Models and analysis for multivariate failure time data* 1992.
[62] J.D. Singer, J.B. Willett, and J.B. Willett, et al. *Applied Longitudinal Data Analysis: modeling change and event occurrence*, Oxford University Press, 2003.
[63] P. Soret, M. Avalos, L. Wittkop, D. Commenges, and R. Thiébaut, *Lasso regularization for left-censored gaussian outcome and high-dimensional predictors*, BMC. Med. Res. Methodol. 18 (2018), pp. 1–13.
[64] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. Van Der Linde, *Bayesian measures of model complexity and fit*, J. R. Stat. Soc. B Stat. Methodol. 64 (2002), pp. 583–639.
[65] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, *Winbugs user manual* 2003.
[66] T.M. Therneau and P.M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer, 2000.
[67] J. Tobin, *Estimation of relationships for limited dependent variables*, Econom. J. Econom. Soc. 26 (1958), pp. 24–36.
[68] J.W. Vaupel, *How change in age-specific mortality affects life expectancy*, Popul. Stud. (Camb). 40 (1986), pp. 147–157.
[69] H.J. Wang, Z. Zhu, and J. Zhou, *Quantile regression in partially linear varying coefficient models*, Ann. Stat. 37 (2009), pp. 3841–3866.
[70] W. Weibull, *Wide applicability*, J. Appl. Mech. 103 (1951), pp. 293–297.
[71] R.E. Wiegand, C.E. Rose, and J.M. Karon, *Comparison of models for analyzing two-group, cross-sectional data with a gaussian outcome subject to a detection limit*, Stat. Methods. Med. Res. 25 (2016), pp. 2733–2749.

Appendices

Appendix 1 – Stellar Dataset

The columns of the dataset in Table 1 are: star name; Type = 1 indicates planet-hosting stars and Type = 2 is the control sample; T_{eff} (in degrees Kelvin) is the stellar surface temperature; log N(Be), log of the abundance of beryllium scaled to the Sun's abundance (i.e. the Sun has log N(Be) = 0.0); log N(Li), log of the abundance of lithium scaled to the Sun's abundance. The indicator variables of left-censoring are given by $\delta_j = 1$ if T is a complete observation and $\delta_j = 0$ if T is a left censored observation, j = 1 (Be) and j = 2 (Li).

Table A1. Stellar astronomy data set.

Row	Star	Type	T_{eff}	δ_1	log[N(Be)]	δ_2	log[N(Li)]
1	HD-6434	1	5835	1	1.08	0	0.80
2	HD-9826	1	6212	1	1.05	1	2.55
3	HD-10647	1	6143	1	1.19	1	2.80
4	HD-10697	1	5641	1	1.31	1	1.96
5	HD-12661	1	5702	1	1.13	0	0.98
6	HD-13445	1	5613	0	0.40	0	−0.12
7	HD-16141	1	5801	1	1.17	1	1.11
8	HD-17051	1	6252	1	1.03	1	2.66
9	HD-19994	1	6109	1	0.93	1	1.99
10	HD-22049	1	5073	1	0.77	0	0.25
11	HD-27442	1	4825	0	0.30	0	−0.47
12	HD-38529	1	5674	0	−0.10	0	0.61
13	HD-46375	1	5268	0	0.80	0	−0.02
14	HD-52265	1	6103	1	1.25	1	2.88
15	HD-75289	1	6143	1	1.36	1	2.85
16	HD-82943	1	6016	1	1.27	1	2.51
17	HD-92799	1	5821	1	1.19	1	1.34
18	HD-95128	1	5924	1	1.23	1	1.83
19	HD-108147	1	6248	1	0.99	1	2.33
20	HD-114762	1	5884	1	0.82	1	2.20
21	HD-117176	1	5560	1	0.86	1	1.88
22	HD-121504	1	6075	1	1.33	1	2.65
23	HD-130322	1	5392	1	0.95	0	0.13
24	HD-134987	1	5776	1	1.22	0	0.74
25	HD-143761	1	5853	1	1.11	1	1.46
26	HD-145675	1	5311	0	0.65	0	0.03
27	HD-169830	1	6299	0	−0.40	0	1.16
28	HD-179949	1	6260	1	1.08	1	2.65
29	HD-187123	1	5845	1	1.08	1	1.21
30	HD-192263	1	4947	0	0.90	0	−0.39
31	HD-195019	1	5842	1	1.15	1	1.47
32	HD-202206	1	5752	1	1.04	1	1.04
33	HD-209458	1	6117	1	1.24	1	2.70
34	HD-210277	1	5532	1	0.91	0	0.30
35	HD-217014	1	5804	1	1.02	1	1.30
36	HD-217107	1	5646	1	0.96	0	0.40
37	HD-222582	1	5843	1	1.14	0	0.59
38	HD-870	2	5447	1	0.80	0	0.20
39	HD-1461	2	5768	1	1.14	0	0.51
40	HD-1581	2	5956	1	1.15	1	2.37
41	HD-3823	2	5948	1	1.02	1	2.41
42	HD-4391	2	5878	1	0.75	0	1.09
43	HD-7570	2	6140	1	1.17	1	2.91
44	HD-10700	2	5344	1	0.83	0	0.41
45	HD-14412	2	5368	1	0.80	0	0.44
46	HD-20010	2	6275	1	1.01	1	2.13
47	HD-20766	2	5733	0	−0.09	0	0.97
48	HD-20794	2	5444	1	0.91	0	0.52
49	HD-20807	2	5843	1	0.36	0	1.07
50	HD-23249	2	5074	0	0.15	1	1.24
51	HD-23484	2	5176	0	0.70	0	0.40
52	HD-26965A	2	5126	1	0.76	0	0.17
53	HD-30495	2	5768	1	1.16	1	2.44
54	HD-36435	2	5479	1	0.99	1	1.67
55	HD-38858	2	5752	1	1.02	1	1.64
56	HD-43162	2	5633	1	1.08	1	2.34
57	HD-43834	2	5594	1	0.94	1	2.30
58	HD-69830	2	5410	1	0.79	0	0.47
59	HD-72673	2	5242	1	0.70	0	0.48
60	HD-74576	2	5000	1	0.70	1	1.72
61	HD-76151	2	5803	1	1.02	1	1.88
62	HD-85117	2	6167	1	1.11	1	2.64
63	HD-189567	2	5765	1	1.06	0	0.82
64	HD-192310	2	5069	0	0.60	0	0.20
65	HD-211415	2	5890	1	1.12	1	1.92
66	HD-222335	2	5260	1	0.66	0	0.31

Appendix 2 – Posterior Bayes Factor (BF)

The posterior Bayes factor is as a discrimination criterion between two models i and j given by $B_{ij} = V_i/V_j$ where V_k is the posterior mean of the likelihood function under model k given by

$$V_k = \int L(D \mid \theta_k)P(\theta_k \mid D)\,\mathrm{d}\theta_k$$

where $L(D \mid \theta_k)$ is the likelihood function under model k and $P(\theta_k \mid D)$ is the joint posterior distribution of the vector of parameters θ_k. If $B_ij = V_i/V_j > 1$, then the Bayes factor criterion favors model i.