Submitted to the Brazilian Journal of Probability and Statistics
arXiv: arXiv:0000.0000

Residual analysis for Weibull accelerated failure time models with random effects[1]

E. S. Rodrigues[a,2], **D. M. Valença**[b] **and J. M. Singer**[c]

[a] *Universidade de São Paulo*
[b] *Universidade Federal do Rio Grande do Norte*
[c] *Universidade de São Paulo*

Abstract. We adapt marginal, conditional and random effects residual plots developed for linear mixed models to assess the fit of Weibull accelerated failure time models with random effects in the presence of censored observations. We propose two imputation procedures to replace the unobserved failure times and perform a simulation study with the objective of studying the ability of residual plots to evaluate the model assumptions under increasing proportions of censoring. We illustrate the proposed residual analysis with a data set involving failure of oil wells.

1 Introduction

We consider a retrospective study related to the operating times of oil wells during the period of 2000 - 2006 designed to identify wells needing preventive maintenance based on some characteristics like production level, lifting method, pump depth, well age, region etc. Since each well may have recurrent failure times, we expect a dependence between the repeated observations on the same well. Moreover, as some wells are disabled from production and others are operating at the end of the study, it is necessary to take censoring into account. Standard parametric accelerated failure time (AFT) models are often used to model data with this nature when the observations are independent [see Lawless (2003), for example]. However, these models are not appropriate to fit correlated survival times. Some authors deal with correlated survival data in the context of reliability of repairable systems [see Ascher and Feingold (1984), Lawless and Thiagarajah (1996) and Percy and Alkali (2007), for example]. Keiding et al. (1997), Lambert et al. (2004), Bolfarine and Valença (2005), Santos and Valença (2012) recommend the

[1]The authors gratefully acknowledge Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant 304126/2015-2), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, grant NF 765/2010) and the CAPES-Proex for partial financial support.

[2]Universidade de São Paulo, Brazil

Keywords and phrases. Censored observations, diagnostics, imputation, linear mixed models

use of AFT models with the inclusion of (non-observable) random effects acting multiplicatively on the event times. Under this parametric setting, a common approach is to use Weibull regression models. Lambert et al. (2004) propose the use of empirical Bayes methods to fit such models, while Carvalho et al. (2014) consider a classical approach based on linear mixed models. These authors advocate the use empirical best linear unbiased predictors (EBLUP), for which only the existence of the first two moments of the distribution of the random effects is required. Carvalho et al. (2014) consider an adaptation of the non-parametric imputation methods proposed by Ageel (2002) to deal with censored data.

For any statistical model, the validity of the underlying assumptions need to be checked by diagnostic techniques such as residual analysis. Nevertheless, there are few proposals for residual analysis in the context of survival models with random effects. Dobson and Henderson (2003), Rizopoulos et al. (2009) and Rizopoulos (2010) define residuals for models that jointly consider the analysis of longitudinal and survival data. These authors, however, assume normality for the distribution of the random effects and do not investigate whether it is valid.

In mixed models, there is more than one source of variability and consequently, more than one type of residuals. Hilden-Minton (1995), Verbeke and Lesaffre (1996a) or Pinheiro and Bates (2000), for example, define three types of residuals to accommodate the additional sources of variability. A summary of the available tools may be found in Nobre and Singer (2007) or Singer et al. (2017), for example.

Our objective is to use residual analysis techniques originally developed for linear mixed models to assess the assumptions on the fixed and random effects employed in correlated survival data models. In this context, the presence of censored observations can distort the interpretation of the residual plots. To bypass this problem we consider imputation methods for the censored data.

In Section 2, we specify the AFT model with random effects, indicate how it may be expressed as a linear mixed model and suggest residual analyses to assess the validity of the associated assumptions. In Section 3, we describe two imputation methods designed to replace the censored observations. In Section 4, we present the results of a simulation study conducted to investigate the impact of different percentages of censored observations on the ability of the proposed residual plots to identify violations of the model assumptions. In Section 5, we analyze the oil well data and we conclude with a brief discussion in Section 6.

2 Accelerated failure time models with random effects

The AFT model with random effects may be written as

$$\ln T_{ij} = b_i + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \sigma\epsilon_{ij}, \tag{2.1}$$

for $i = 1,\ldots,k$, $j = 1,\ldots,n_i$, where T_{ij} represents the time between the $(j-1)$-th and the j-th failure of the i-th sample unit, b_i are independent and identically distributed (unobserved) random effects with null means and common variance σ_b^2, $\mathbf{x}_{ij}$ is a $(p\times 1)$ vector of covariates with the first component equal to 1, $\boldsymbol{\beta}$ is a $(p\times 1)$ vector whose elements are fixed (but unknown) parameters, σ is a (unknown) scale parameter and ϵ_{ij} are independent and identically distributed unobserved random errors with known mean and known common variance σ_ϵ^2. Furthermore, we assume that $Cov(b_i, \epsilon_{ij}) = 0$. When $\sigma_b^2 = 0$, this model reduces to the usual AFT model [see Lawless (2003) and Bolfarine and Valença (2005), for example].

Because of censoring, the response variable $\ln T_{ij}$ is not observed for all sampling units. In fact, we observe

$$Y_{ij} = \delta_{ij} \ln T_{ij} + (1-\delta_{ij}) \ln C_{ij}, \tag{2.2}$$

where C_{ij} is the j-th censoring time for the i-th sample unit and $\delta_{ij} = I(T_{ij} \le C_{ij})$ is an indicator of failures.

Consider initially a sample without censoring, *i.e.*, suppose that $\delta_{ij} = 1$ for $i = 1,\ldots,k$, $j = 1,\ldots,n_i$ in (2.2) and let $n = \sum_{i=1}^k n_i$.

Then the AFT model with random effects (2.1) can be represented as a linear mixed model, namely

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}, \tag{2.3}$$

where $\mathbf{Y} = (\mathbf{Y}_1^\top,\ldots,\mathbf{Y}_k^\top)^\top$, with $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{in_i})^\top$, $i = 1,\ldots,k$, $\mathbf{X} = (\mathbf{X}_1^\top,\ldots,\mathbf{X}_k^\top)^\top$, $\mathbf{X}_i = (\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{in_i})^\top$, $\mathbf{Z} = \oplus_{i=1}^k \mathbf{1}_{n_i}$, $\mathbf{b} = (b_1,\ldots,b_k)^\top$ is a $(k\times 1)$ vector of random effects such that $E(\mathbf{b}) = \mathbf{0}$, $Var(\mathbf{b}) = \sigma_b^2 \mathbf{I}_k$,

$$\mathbf{e} = \sigma[\boldsymbol{\epsilon} - E(\boldsymbol{\epsilon})]$$

is an $(n\times 1)$ vector of random uncorrelated errors and $\boldsymbol{\epsilon}$ is a $(n\times 1)$ vector having the random errors ϵ_{ij} as components. Thus, $E(\mathbf{e}) = \mathbf{0}$, $Var(\mathbf{e}) = \sigma_e^2 \mathbf{I}_n$, with $\sigma_e^2 = \sigma^2 Var(\epsilon_{ij})$. When ϵ_{ij} follows a standard extreme value distribution, *i.e.*, $\exp(\epsilon_{ij})$ follows a Weibull distribution, then $\sigma_e^2 = \sigma^2\pi^2/6$. These definitions imply that

$$\begin{aligned} E(\mathbf{Y}) &= \mathbf{X}\boldsymbol{\beta} \\ Var(\mathbf{Y}) &= \mathbf{V} = \sigma_b^2 \mathbf{Z}\mathbf{Z}^\top + \sigma_e^2 \mathbf{I}_n \\ Cov(\mathbf{b},\mathbf{Y}^\top) &= \mathbf{C} = \sigma_b^2 \mathbf{Z}^\top. \end{aligned} \tag{2.4}$$

Conditionally on the knowledge of $\mathbf{V}$ and $\mathbf{C}$, the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) of $\mathbf{b}$ are respectively:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{Y} \quad \text{and} \quad \tilde{\mathbf{b}} = \mathbf{C}\mathbf{V}^{-1}(\mathbf{Y}^{-}\mathbf{X}\widehat{\boldsymbol{\beta}}). \tag{2.5}$$

In practice, as the variance components σ_b^2 and σ_e^2 (and therefore, $\mathbf{V}$ and $\mathbf{C}$) are unknown, it is reasonable to use empirical estimates and predictors (EBLUE and EBLUP, respectively), obtained by replacing $\mathbf{V}$ and $\mathbf{C}$ with suitable estimators. Details can be found in Robinson (1991) and Jiang and Verbeke (1998). Estimates of the variance components may be obtained via the nonparametric MINQUE and I-MINQUE methods [see Rao (1971a), Rao (1971b) and Searle et al. (1992), for example].

Under model (2.3) we may consider three types of residuals, namely:

- **Marginal residuals**: $\widehat{\boldsymbol{\xi}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, that predict the marginal errors, $\boldsymbol{\xi} = \mathbf{Y} - E(\mathbf{Y}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\mathbf{b} + \mathbf{e}$
- **Conditional residuals**: $\widehat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\tilde{\mathbf{b}}$, that predict the conditional errors, $\mathbf{e} = \mathbf{Y} - E(\mathbf{Y}|\mathbf{b}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$
- **Random effects residuals**: $\mathbf{Z}\tilde{\mathbf{b}}$, that predict the random effects, $\mathbf{Z}\mathbf{b} = E(\mathbf{Y}|\mathbf{b}) - E(\mathbf{Y}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) - (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

The marginal residuals may be used to evaluate the linearity of fixed effects, to detect outliers as well as to check the adequacy of the within-unit covariance structure. Lesaffre and Verbeke (1998) comment that when the within-unit covariance structure is adequate, $\mathcal{V}_i = \|\mathbf{I}_{n_i} - \widehat{\mathcal{R}}_i\widehat{\mathcal{R}}_i^\top\|^2$, $i = 1,\ldots,k$, where $\widehat{\mathcal{R}}_i = \widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{-1/2}\widehat{\boldsymbol{\xi}}_i$ with $\widehat{\boldsymbol{V}}_i = \boldsymbol{V}_i(\widehat{\theta})$ being the i-th diagonal block of $\boldsymbol{V}$, should be close to zero. Units with large values of $\mathcal{V}_i$ are those for which the proposed covariance structure might not be adequate. Given that the true variance of $\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}}$ is $\mathbb{V}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}}) = [\boldsymbol{V}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}(\boldsymbol{X}_{\boldsymbol{i}}^\top\boldsymbol{V}_{\boldsymbol{i}}^{-1}\boldsymbol{X}_{\boldsymbol{i}})^{-1}\boldsymbol{X}_{\boldsymbol{i}}^\top]$ and not $\boldsymbol{V}_{\boldsymbol{i}}$, Singer et al. (2017) consider replacing $\widehat{\mathcal{R}}_i$ in $\mathcal{V}_i$ with $\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}}^* = [\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}})]^{-1/2}\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}}$ where $\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}})$ corresponds to the diagonal block of $\widehat{\boldsymbol{V}} - \boldsymbol{X}(\boldsymbol{X}^\top\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ associated to the i-th unit. Furthermore, to avoid giving much weight to units with many observations, these authors consider taking $\mathcal{V}_i^* = \sqrt{\mathcal{V}_i}/n_i$ as a standardized measure of adequacy of the within-unit covariance structure. Plots of $\mathcal{V}_i^*$ as functions of the unit indices, i, (termed unit index-plots) may help to identify units for which the covariance structure should be modified.

To evaluate the linearity of the fixed effects in model (2.3), Singer et al. (2017) consider plotting the elements of the standardized marginal residuals $\widehat{\xi}_{ij}^* = \widehat{\xi}_{ij}/[diag_j(\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}}))]^{1/2}$, where $diag_j(\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}}))$ is the j-th element of the main diagonal of $\widehat{\mathbb{V}}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{i}})$, *versus* the values of each explanatory variable as well as *versus* the fitted values. They also recommend plotting $\widehat{\xi}^*_{ij}$ *versus* the observation index as a tool to detect outlying observations.

Given that $\mathbb{V}(\widehat{\mathbf{e}}) = \sigma_e^4[\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{V}^{-1}]$, Nobre and Singer (2007) observe that the conditional residuals may have different variances. They suggest plots of standardized conditional residuals, $\widehat{e}^*_{ij} = \widehat{e}_{ij}/diag_{ij}(\widehat{\mathbb{V}}(\widehat{\mathbf{e}}))^{1/2}$, with $diag_{ij}(\widehat{\mathbb{V}}(\widehat{\mathbf{e}}))$ denoting the main diagonal element of $\widehat{\mathbb{V}}(\widehat{\mathbf{e}})$ corresponding to the j-th observation of the i-th unit *versus* fitted values to check for homoskedasticity of the conditional errors or *versus* unit index to check for outlying observations.

When there is no confounding and the random effects follow a q-dimensional gaussian distribution, $\mathcal{M}_i = \tilde{\mathbf{b}}_i^\top\{\widehat{\mathbb{V}}[\tilde{\mathbf{b}}_i - \mathbf{b}_i]\}^{-1}\tilde{\mathbf{b}}_i$ (the Mahalanobis's distance between $\tilde{\mathbf{b}}_i$ and $\mathbb{E}(\mathbf{b}_i) = \mathbf{0}$) should have an approximate chi-squared distribution with q degrees of freedom. Therefore, a χ^2_q QQ plot for $\mathcal{M}_i$ may be used to verify whether the random effects follow a (q-variate) gaussian distribution. Unit index-plots of $\mathcal{M}_i$ may also be employed to detect outliers. More details on residual analysis for mixed models may be obtained in Singer et al. (2017).

3 Taking censored observations into account

The response variables (2.2) underestimate the true times between failures when $\delta_{ij} = 0$ so that an appropriate use of the standard linear mixed model (2.3) requires some form of imputation. In this context, we propose two procedures to replace the censored values C_{ij} in (2.2) by estimates $\widehat{T}_{ij}$ of the true (unobserved) failure time T_{ij}, *i.e.*, we consider the response variable

$$Y^*_{ij} = \delta_{ij}\ln T_{ij} + (1-\delta_{ij})\ln\widehat{T}_{ij}, \tag{3.1}$$

for $i = 1, \dots, k$ and $j = 1, \dots, n_i$.

3.1 Extension of Ageel's method (EAM)

Ageel (2002) proposes parametric and non-parametric methods to impute the censored observations in a survival model with independent random and right censored observations. Carvalho et al. (2014) consider an adaptation of Ageel's nonparametric approach to deal with correlated data. Here, we adapt the parametric approach to incorporate covariates as in (2.1) assuming that conditionally to random effects b_i, T_{ij} follows a two parameter Weibull distribution and replacing each censored value with an estimate of $E(T_{ij}|T_{ij} > c_{ij}, b_i, \mathbf{x}_{ij})$.

The density and survival functions of the conditional distribution of T_{ij} given b_i are, respectively:

$$f(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij}) = \frac{\gamma}{\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})} \left(\frac{t_{ij}}{\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})} \right)^{\gamma - 1} \times$$
$$\times \exp\left[- \left(\frac{t_{ij}}{\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})} \right)^{\gamma} \right] I_{(0,\infty)}(t_{ij})$$

and

$$S(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij}) = \exp\left[- \left(\frac{t_{ij}}{\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})} \right)^{\gamma} \right] I_{(0,\infty)}(t_{ij}),$$

where $\boldsymbol{\phi} = (\boldsymbol{\beta}^\top, \sigma^2, \sigma_b^2)^\top$ is the vector of parameters, $\gamma = 1/\sigma > 0$ denotes the shape parameter and $\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x_{ij}}) > 0$ represents the scale parameter. The density function and the expected value for the conditional distribution of T_{ij} (given $T_{ij} > c_{ij}$ and the random effects vector b_i) are, respectively

$$f(t_{ij}|t_{ij} > c_{ij}; b_i, \boldsymbol{\phi}, \mathbf{x}_{ij}) = \frac{f(t_{ij}|b_i, \mathbf{x}_{ij})}{S(c_{ij}|b_i, \mathbf{x}_{ij})} =$$
$$= \frac{\gamma/[\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})] \left(t_{ij}/[\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})] \right)^{\gamma-1}}{\exp\left[- \left(c_{ij}/\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij}) \right)^{\gamma} \right]} \times$$
$$\times \frac{\exp\left[- \left(t_{ij}/[\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})] \right)^{\gamma} \right]}{\exp\left[- \left(c_{ij}/\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij}) \right)^{\gamma} \right]}$$

and

$$E(T_{ij}|T_{ij} > c_{ij}; b_i, \boldsymbol{\phi}, \mathbf{x}_{ij}) = \frac{\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij}) \Gamma\left(1 + 1/\gamma, \left(c_{ij}/\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij}) \right)^{\gamma} \right)}{\exp\left[- \left(c_{ij}/\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij}) \right)^{\gamma} \right]}, \tag{3.2}$$

where c_{ij} represents the time of occurrence of the j-th censoring for the i-th sample unit and $\Gamma(\eta,\tau)$ denotes the incomplete gamma function with parameters $\eta = 1 + 1/\gamma$ and $\tau = \left(c_{ij}/\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}) \right)^{\gamma}$ [Olver et al. (2010)].

3.2 Inverse transform method (ITM)

A second imputation approach consists in replacing the censored observations in (2.2) by a value generated from the conditional distribution of T_{ij} given the random effect b_i and $T_{ij} > c_{ij}$. We generate random observations $\widehat{T}_{ij}$ from the cumulative distribution function $F(t_{ij}|T_{ij} > c_{ij}, b_i, \mathbf{x}_{ij})$ using the method of inverse transform [see, for example, Stewart (2009)] and plug in the generated values in (3.1).

The distribution function of T_{ij} given the vector of random effects b_i, the vector of covariates $\mathbf{x}_{ij}$ and $T_{ij} > c_{ij}$ is

$$F(t_{ij}|t_{ij} > c_{ij}; b_i, \boldsymbol{\phi}, \mathbf{x}_{ij}) = 1 - \frac{S(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij})}{S(c_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij})}. \tag{3.3}$$

We obtain t_{ij} through the inverse of the distribution function given in (3.3), and thus, $\widehat{T}_{ij}$ in (3.1) can be generated from

$$\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij}) \left[-\log(1 - u_{ij}) + \left(\frac{c_{ij}}{\exp(b_i + \boldsymbol{\beta}^\top \mathbf{x}_{ij})} \right)^{\gamma} \right]^{1/\gamma}, \tag{3.4}$$

where u_{ij} is a value generated from a uniform distribution on (0,1), $i = 1, \ldots, k$, $j = 1, \ldots, n_i$.

3.3 Procedures for imputation

To obtain estimates of (3.2) and (3.4) we require preliminary estimates of $\boldsymbol{\phi}$ and of the predictors of b_i which may be obtained from the penalized log-likelihood [see Therneau et al. (2003)]

$$l_{pen}(\boldsymbol{\phi}|\boldsymbol{b}) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log f(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij})^{\delta_{ij}} S(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij})^{1-\delta_{ij}} - h(\boldsymbol{b}; \sigma_b^2), \tag{3.5}$$

where $f(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij})$ and $S(t_{ij}|b_i, \boldsymbol{\phi}, \mathbf{x}_{ij})$ are, respectively, the density and survival functions of the conditional distribution of T_{ij} given b_i and $h(\boldsymbol{b}; \sigma_b^2)$ is the penalty function. When the distribution of the random effects is normal, Therneau et al. (2003) suggest the following penalty function

$$h(\boldsymbol{b}; \sigma_b^2) = \frac{1}{2\sigma_b^2} \sum_{i=1}^{k} b_i^2. \tag{3.6}$$

Maximizing (3.5), we obtain the maximum penalized likelihood estimator $\widehat{\boldsymbol{\phi}}^*$ of $\boldsymbol{\phi}$ and predictors $\tilde{b}_i^*$ of the random effects b_i, $i = 1, \ldots, k$. In this context we may use the function `survreg` in the library `survival` available in the free software package `R` R Development Core Team (2015). Details can be found in Santos and Valença (2012).

The procedure for imputing the censored data and obtaining the residuals may be summarized as follows.

i) Assume that $b_i \sim N(0, \sigma_b^2)$ in (2.1);
ii) obtain $\widehat{\boldsymbol{\phi}}^*$ and $\tilde{b}_i^*$ based on the penalized likelihood (3.5);

iii) select the significant covariates using likelihood ratio tests;
iv) compute Y_{ij}^* in (3.1), considering for each censored value, an estimate $\widehat{T}_{ij}$ from either (3.2) or (3.4) with $\boldsymbol{\phi}$ and b_i replaced by $\widehat{\boldsymbol{\phi}}^*$ and $\tilde{b}_i^*$ and u_{ij} generated from a standard uniform distribution on (0,1);
v) repeat step iv) for all censored values and obtain the vector $\mathbf{Y}^* = (\mathbf{Y}_1^{*\top}, \ldots, \mathbf{Y}_k^{*\top})^\top$ where $\mathbf{Y}_i^* = (Y_{i1}^*, \ldots, Y_{in_i}^*)^\top$, $i = 1, \ldots, k$.

After replacing the censored observations with the imputed values, we may treat the vector $\mathbf{Y}^*$ as the true vector of responses in (2.3). The estimates of the parameters (EBLUE) and predictors of the random effects (EBLUP) do not require a specification for the distribution of the random effects.

4 Simulation

A simulation study was conducted to evaluate the performance of the residuals plots developed for linear mixed models, in AFT models with random effects when censored observations are imputed via each of the two proposed procedures, **EAM** and **ITM**.

Specifically, we considered the model (2.1) with a single covariate x_{ij} drawn from the standard normal distribution. In order to mimic the structure of the data motivating this study, we considered samples with $k = 200$ units, with sizes n_i for $i = 1, ..., k$, generated from a *Poisson* distribution with mean 6, where units with size zero were disregarded. The true parameter values were taken as $\beta_0 = 7.5$, $\beta_1 = 0.85$ and $\sigma = 1.15$. The random effects b_i were generated as independent and identically distributed ($i.i.d$) normal random variables with zero mean and variance $\sigma_b^2 = 0.35$ and the errors ϵ_{ij} were generated as $i.i.d$ standard extreme value random variables. This yields failure times t_{ij} that, conditionally on the random effects b_i, follow a Weibull distribution. We fitted the model to each sample considering $\epsilon_i \sim VEP$ and no assumption on form of the distribution of the random effects b_i.

Besides uncensored samples, we considered samples with 30% and 50% censored observations. Censoring times c_{ij} were generated as independent uniform random variables on $(0, u)$, where u varies according to the specified censoring proportions. Censored values were replaced by data obtained from each imputation method and model (2.3) with response variable (3.1) was fitted to the data.

The model assumptions were checked via the following plots:

(i) Modified Lesaffre-Verbeke index ($\mathcal{V}_i^*$) unit index plots;
(ii) Standardized marginal residuals ($\hat{\boldsymbol{\xi}}_i^*$) vs fitted values;
(iii) Standardized conditional residuals ($\hat{e}_k^*$) vs fitted values;

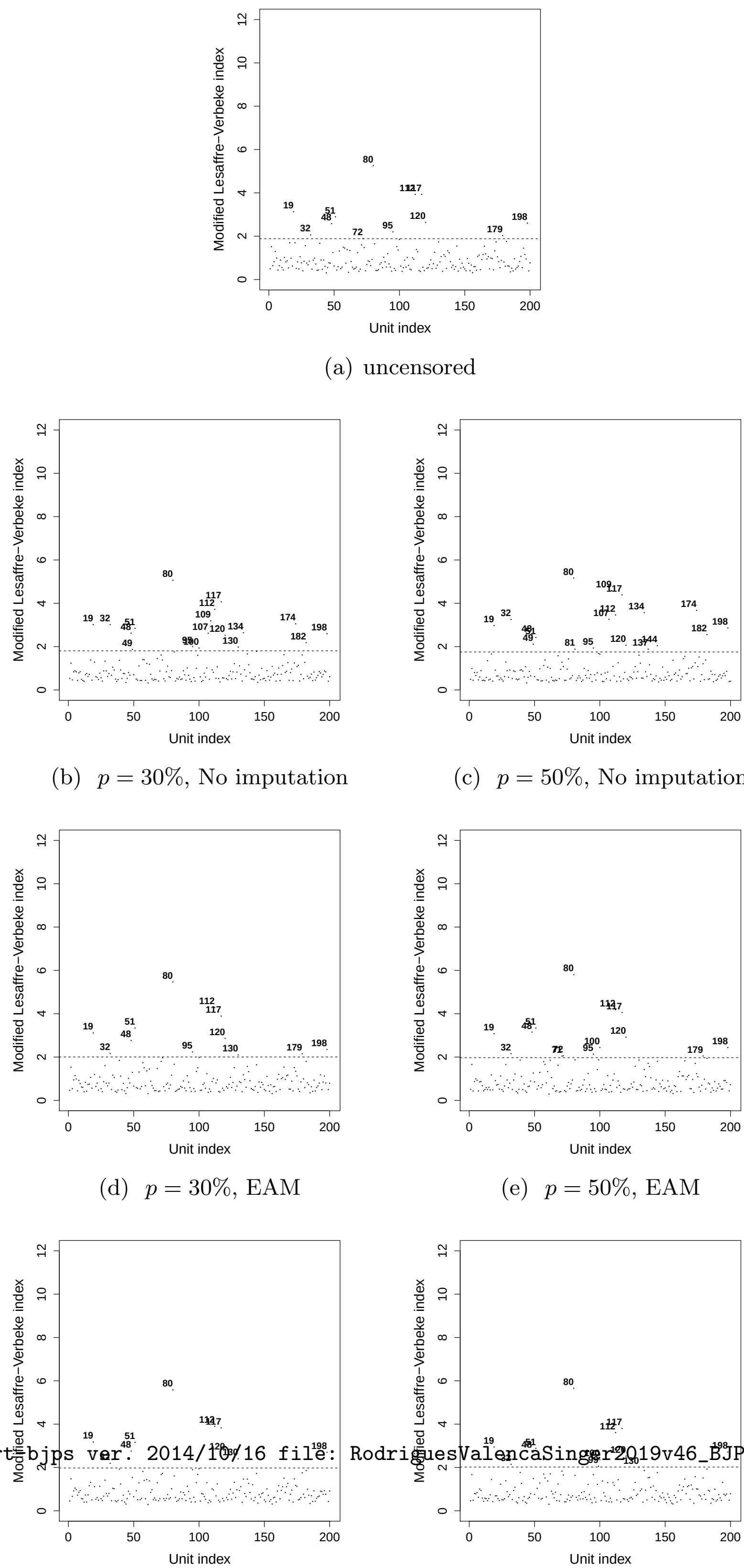
(iv) QQ plot of conditional residuals ($\mathbf{e}$);
(v) QQ plot of Mahalanobis' distance ($\mathcal{M}_i$).

To evaluate the performance of the two imputation procedures, we simulated 1,000 data sets for each of three scenarios: uncensored observations, 30% and 50% censored observations.

To illustrate the diagnostic procedure, we selected typical samples obtained with complete, 30% and 50% censored observations as well as with the imputed data; the corresponding residual plots are displayed in (Figures 1-5). In general, residual plots obtained for the censored data suggest that the model assumptions are violated but this is corrected or at least attenuated when the plots obtained with the imputed data are considered. The Modified Lesaffre-Verbeke index plots displayed in Figure 1 are exceptions, since they show the same pattern for uncensored, censored or imputed data, suggesting that the adopted covariance structure is adequate for around 95% of the units.

The standardized marginal residual plots obtained with censored data [Figures 2(b) and 2(c)], on the other hand, suggest that the linearity assumption may not hold, with increasing evidence for the heavier censored scenario. In these cases, the plots obtained from the **ITM** imputed data [Figures 2(d) and 2(f)] indicate that this procedure restores the original uncensored data pattern [see Figure 2(a)]. Similar conclusions may be derived from the inspection of the standardized conditional residual plots; those obtained under the censored scenarios [see Figures 3(b) and 3(c)] suggest a violation of the homoscedasticity assumption for the conditional errors.

The QQ plots for conditional residuals suggest a slight violation of the normality assumption for the conditional errors specially for 50% censoring [see Figure 4(c)], which seems to be reasonably corrected with imputation via **ITM** [Figure 4(g)]. A similar conclusion holds for the Mahalanobis distance QQ plots (Figure 5), but both imputation methods seem to restore the pattern obtained with the complete data. Note that even if such plots detect violations of the normality assumption, this does not have implications for the models we are considering, given that they do not require a form for the distribution of the random effects.



(a) uncensored

(b) $p = 30\%$, No imputation

(c) $p = 50\%$, No imputation

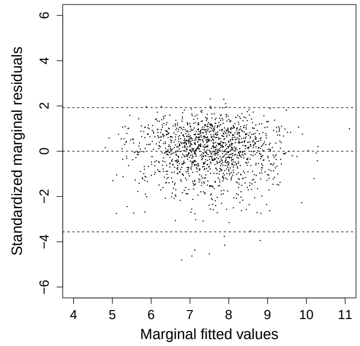
(d) $p = 30\%$, EAM

(e) $p = 50\%$, EAM

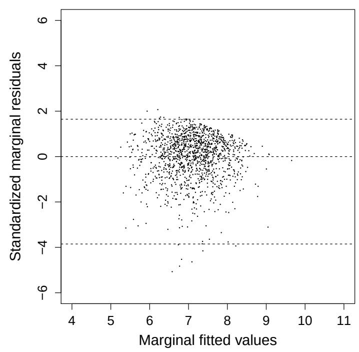
(f) $p = 30\%$, ITM

(g) $p = 50\%$, ITM

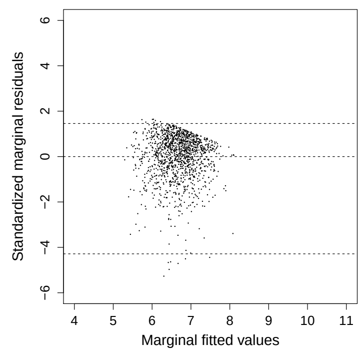
Figure 1 *Modified Lesaffre-Verbeke index ($\mathcal{V}_i^*$) unit index plots.*



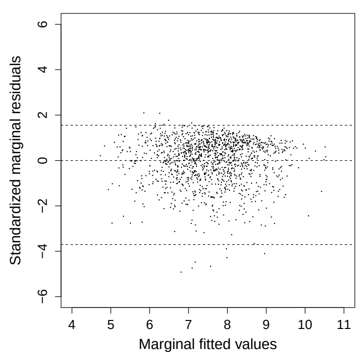
(a) uncensored



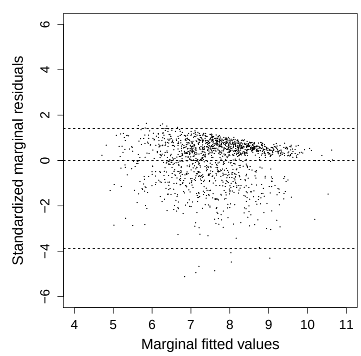
(b) $p = 30\%$, No imputation



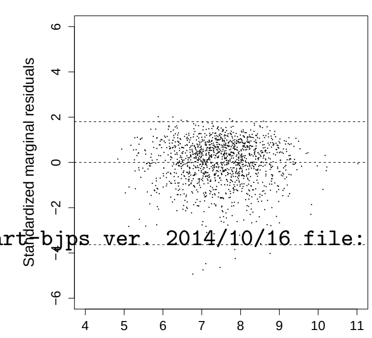
(c) $p = 50\%$, No imputation



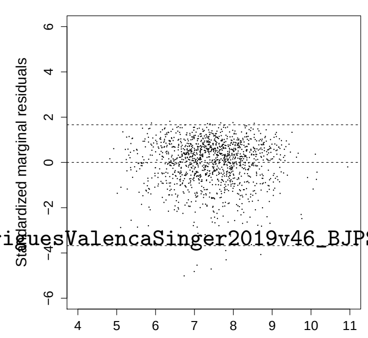
(d) $p = 30\%$, EAM



(e) $p = 50\%$, EAM

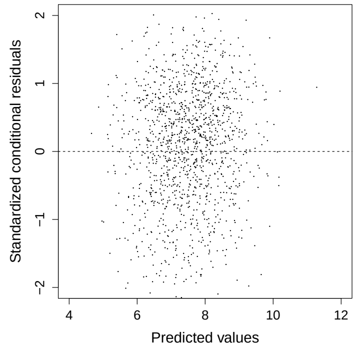


(f) $p = 30\%$, ITM

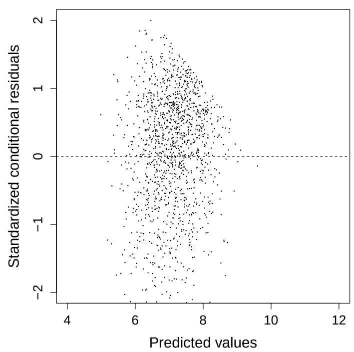


(g) $p = 50\%$, ITM

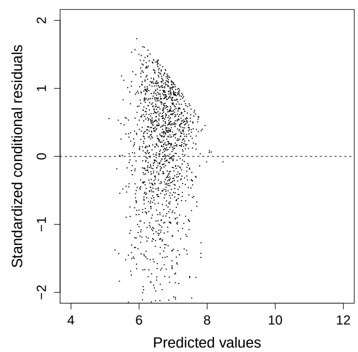
Figure 2 *Standardized marginal residuals ($\hat{\boldsymbol{\xi}}_i^*$) vs fitted values.*



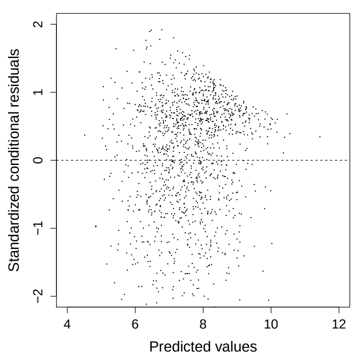
(a) uncensored



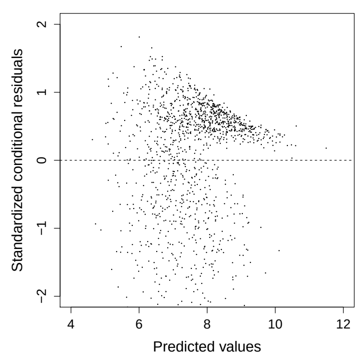
(b) $p = 30\%$, No imputation



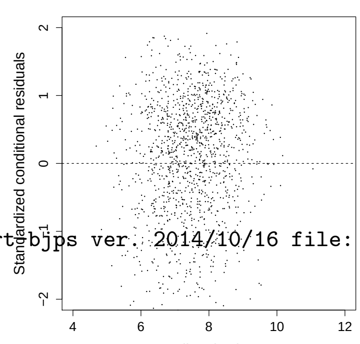
(c) $p = 50\%$, No imputation



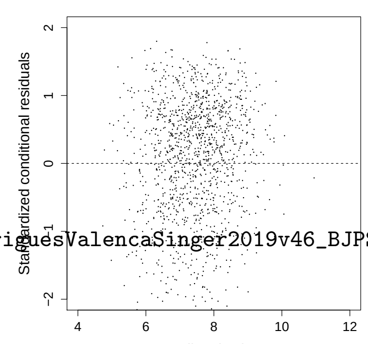
(d) $p = 30\%$, EAM



(e) $p = 50\%$, EAM

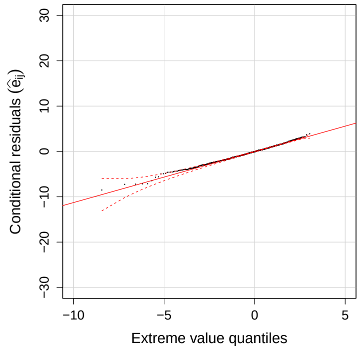


(f) $p = 30\%$, ITM

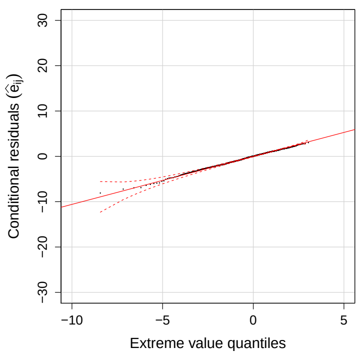


(g) $p = 50\%$, ITM

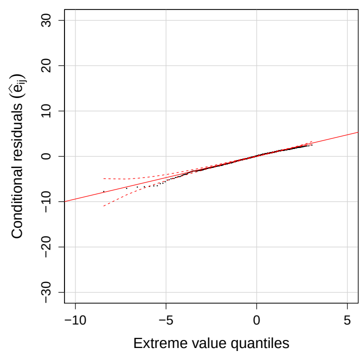
Figure 3 *Standardized conditional residuals ($\hat{e}_k^*$) vs fitted values.*



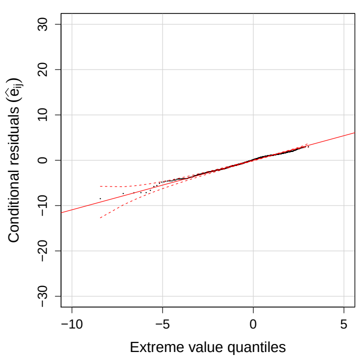
(a) uncensored



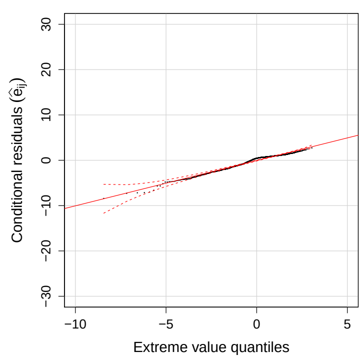
(b) $p = 30\%$, No imputation



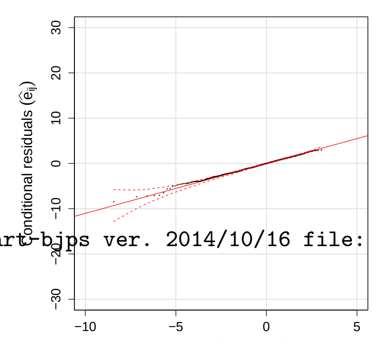
(c) $p = 50\%$, No imputation



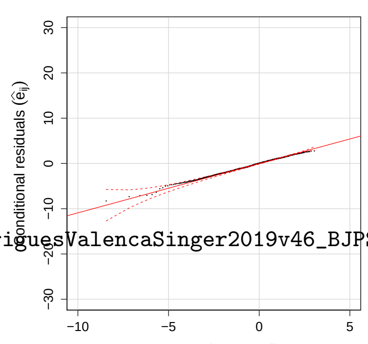
(d) $p = 30\%$, EAM



(e) $p = 50\%$, EAM

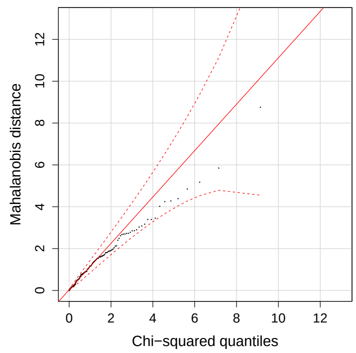


(f) $p = 30\%$, ITM

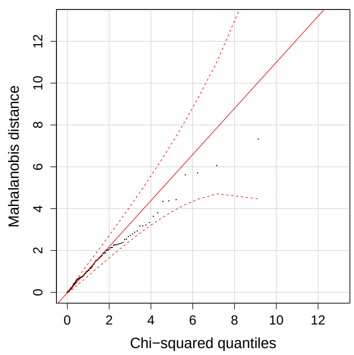


(g) $p = 50\%$, ITM

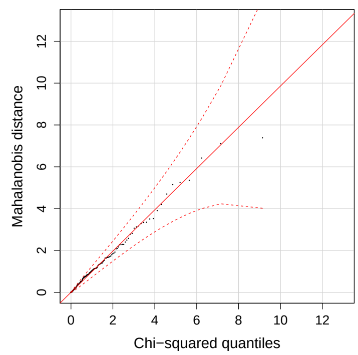
Figure 4 *QQ plot for conditional residual* (**e**).



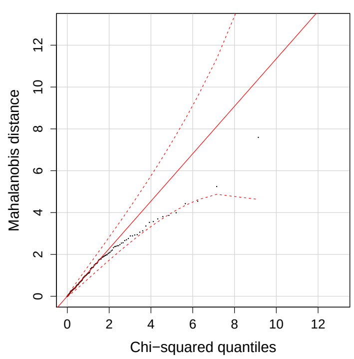
(a) uncensored



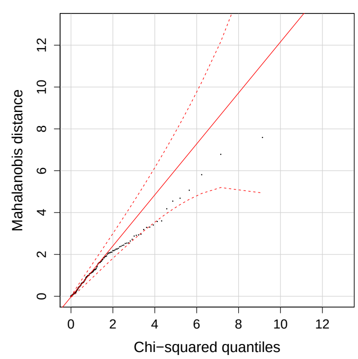
(b) $p = 30\%$, No imputation



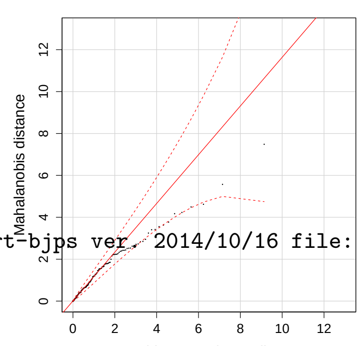
(c) $p = 50\%$, No imputation



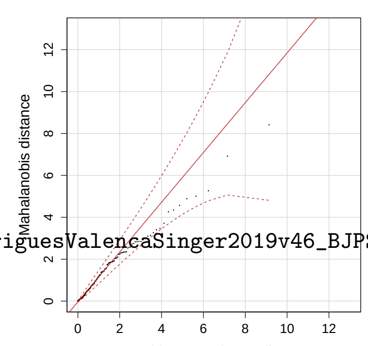
(d) $p = 30\%$, EAM



(e) $p = 50\%$, EAM



(f) $p = 30\%$, ITM



(g) $p = 50\%$, ITM

Figure 5 *QQ for Mahalanobis distance ($\mathcal{M}_i$).*

An additional procedure to evaluate the similarity between the results obtained via imputed data and those obtained with complete data may be conducted via the comparison of the distribution functions of the corresponding conditional residuals. In this context we may employ CramÃl'r-von Mises or Kolmogorov-Smirnov statistics among other alternatives. Letting Q_n denote the empirical distribution function of the censored, **ITM** or **EAM** imputed data and P_n denote the empirical distribution function of the complete data, the two sample Cramér-von Mises statistic is

$$CM(Q_n, P_n) = \frac{U}{2n^3} - \frac{4n^2 - 1}{12n}$$

with

$$U = n\{\sum_{i=1}^{n}(r_i - i)^2 + \sum_{j=1}^{n}(s_j - j)^2\}$$

where r_i (s_j) denotes the rank of the i-th (j-th) conditional residuals obtained with the imputed (complete) data in the combined sample. Details are given in Anderson (1962).

A similar evaluation may be based on the two sample Komolgorov-Smirnov statistic

$$KS(Q_n, P_n) = sup_x|Q_n(x) - P_n(x)|.$$

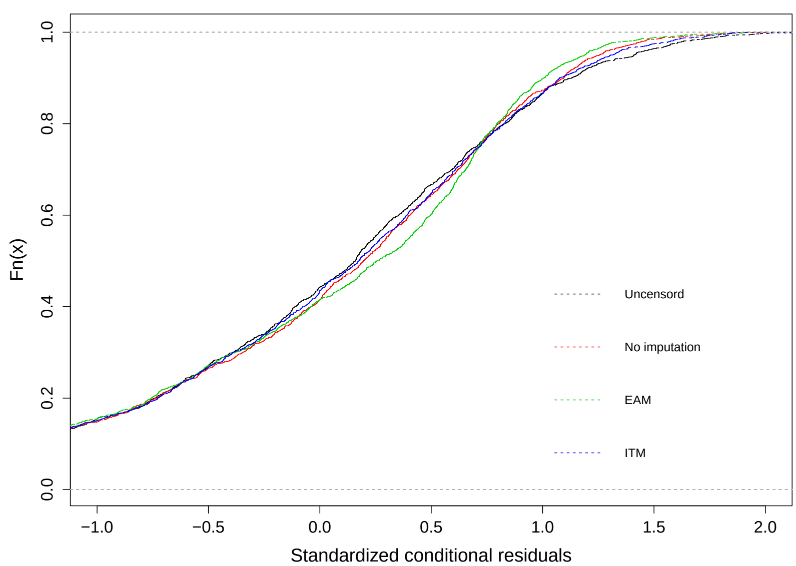
We applied both the CramÃl'r-von Mises and the Kolmogorov-Smirnov statistics to compare the distributions of the conditional residuals obtained with the censored, EAM imputed as well as with the ITM imputed data with those obtained with the complete data for each of the simulated data sets under 30% and 50% censoring schemes. The corresponding minimum, maximum and average p-values as well as the rejection rate of the null hypotheses of equality between the distributions of the conditional residuals corresponding to the nominal level of 5% are displayed in Table 1.

The difference between the empirical distribution functions of the conditional residuals obtained with complete, 30% and 50% censored observations as well as with the imputed data plotted in Figures (6) - (7) for selected samples. The empirical distribution functions generated by the data imputed by the ITM are clearly closer to those obtained from the complete data. Considering that, by Table 1, this is the behaviour for most cases, we understand that this method is more convenient for our purposes.

Table 1 *p-values and null percentage rejection rates (% NRR) for the Cramér-von Mises (CvM) and Kolmogorov-Smirnov (KS) tests*

Censoring percentage	Imputation Method	Test	p-values minimum	maximum	average	% NRR
30%	No	CvM	0.1485	0.9901	0.5637	0.0000
30%	No	KS	0.0233	0.9944	0.5492	0.1000
30%	EAM	CvM	0.0000	0.2178	0.0542	58.000
30%	EAM	KS	0.0001	0.0997	0.0105	98.500
30%	ITM	CvM	0.5149	0.9901	0.9608	0.0000
30%	ITM	KS	0.4208	1.0000	0.9326	0.0000
50%	No	CvM	0.0000	0.4951	0.0929	34.700
50%	No	KS	0.0020	0.6981	0.1316	23.000
50%	EAM	CvM	0.0000	0.0009	0.0000	100.00
50%	EAM	KS	0.0000	0.0000	0.0000	100.00
50%	ITM	CvM	0.1881	0.9901	0.8116	0.0000
50%	ITM	KS	0.0907	1.0000	0.7570	0.0000

Figure 6 *Empirical distribution functions forconditional residuals (30% censorship)*

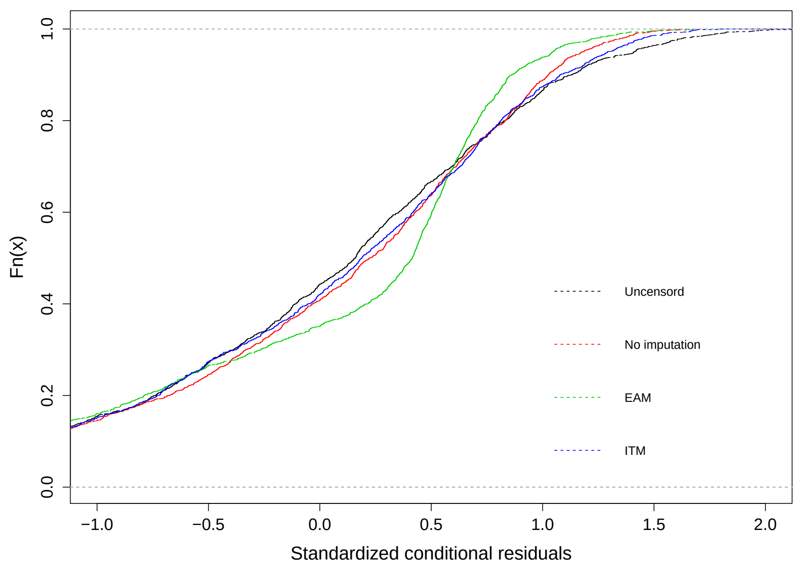


5 Analysis of the oil-well data

5.1 Model Specification

To illustrate the proposed methodology we consider 2374 observations from 616 oil wells involving 1811 failure times and 563 censored times. The failures

Figure 7 *Empirical distribution functions for conditional residuals (50% censorship)*



were characterized by the total interruption of the operation of the well due to malfunction of one or more components of the sub-surface equipment. The covariates were, *Oil production* in m^3/day (PRO), *Elevation method*: [pumpjack ($PJ = 1$) or progressive cavity ($PJ = 0$)], *Age at failure* in *years* (AGE), *Location of operating unit*: [Region B ($RB = 1$ and $RC = RD = 0$), C ($RC = 1$ and $RB = RD = 0$), D ($RD = 1$ and $RB = RC = 0$) and A ($RB = RC = RD = 0$)] and *Depth of the oil pump* in *meters* (DEP).

Following the suggestions of Carvalho et al. (2014) we adopted the model following

$$\begin{aligned}\ln T_{ij} = {} & b_i + \beta_{pro}PRO_{ij} + \beta_{pj}PJ_{ij} + \beta_{age}AGE_{ij} + \beta_{rb}RB + \beta_{rc}RC + \beta_{rd}RD+ \\ & + \beta_{dep}DEP_i + \beta_{pro*rb}PRO_{ij} * RB + \beta_{pro*rc}PRO_{ij} * RC+ \\ & + \beta_{pro*rd}PRO_{ij} * RD + \beta_{dep*rb}DEP_i * RB+ \\ & + \beta_{dep*rc}DEP_i * RC + \beta_{dep*rd}DEP_i * RD + \sigma\epsilon_{ij},\end{aligned} \tag{5.1}$$

assuming that, conditionally on the random effects b_i, the T_{ij}, $i = 1, \ldots, 616$ and $j = 1, \ldots, n_i$ follow Weibull distributions. The b_i are i.i.d. random effects with null means and variance σ_b^2. The form of their distribution is not specified.

Censored observations were imputed via both the ITM and the EAM procedures and the EBLUE of the parameters were computed using (2.5). The variance components were estimated by the MINQUE method. Estimates of the fixed parameters and of the variance components are displayed in Table 2. Both approaches produce comparable estimates of the model parameters.

Table 2 *Estimates and standard errors (SE) for model (5.1) parameters*

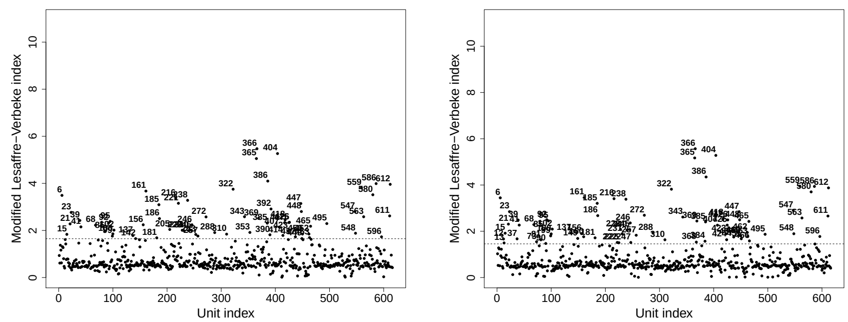
	ITM		EAM	
Parameter	Estimate	SE	Estimate	SE
β_0	6.577	0.302	6.479	0.302
β_{pro}	-0.041	0.012	-0.041	0.012
β_{pj}	0.623	0.152	0.646	0.152
β_{age}	0.088	0.008	0.092	0.008
β_{rb}	1.334	0.371	1.455	0.371
β_{rc}	1.125	0.277	1.177	0.277
β_{rd}	1.804	0.394	1.950	0.394
β_{dep}	0.002	<0.001	0.002	<0.001
β_{pro*rb}	0.042	0.019	0.041	0.019
β_{pro*rc}	0.004	0.019	0.003	0.019
β_{pro*rd}	-0.026	0.024	-0.027	0.024
β_{dep*rb}	-0.002	0.001	-0.002	0.001
β_{dep*rc}	-0.003	0.001	-0.003	0.001
β_{dep*rd}	-0.003	0.001	-0.003	0.001
σ_e	1.451	-	1.440	-
σ_b^2	0.896	-	0.908	-

5.2 Residual analysis of the fitted model

The residual plots corresponding to the fit of model (5.1) are shown in Figures 8- 14. The results suggest that:

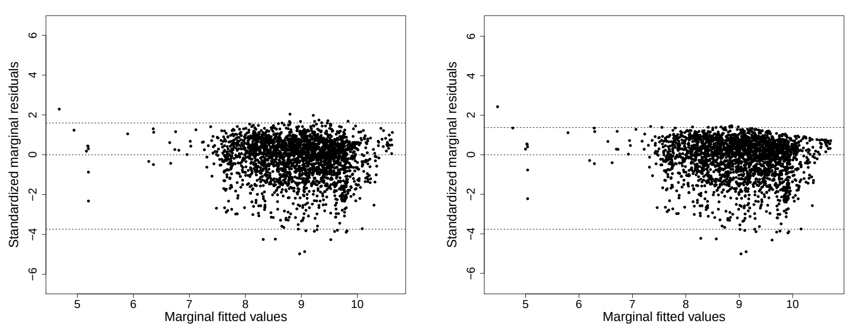
- The structure of the within-units covariance matrix proposal may not be considered adequate only for a small number (13%) of some sample units (Figure 8).
- The hypothesis of linearity of the fixed effects appears satisfactory (Figure 9) and the distribution of the marginal residuals are slightly negatively skewed as expected (Figure 10).
- There is no evidence of outliers (Figure 11).
- There is evidence of violation of the homoskedasticity assumption for the conditional errors (Figure 12).
- The standard extreme value distribution assumption seems acceptable for the conditional errors (Figure 13).

- There is some evidence against normality of the random effects (Figure 14); this however, is not a model requirement.



(a) ITM (b) EAM

Figure 8 *Oil well data - Modified Lesaffre-Verbeke index ($\mathcal{V}_i^*$) unit index plots.*

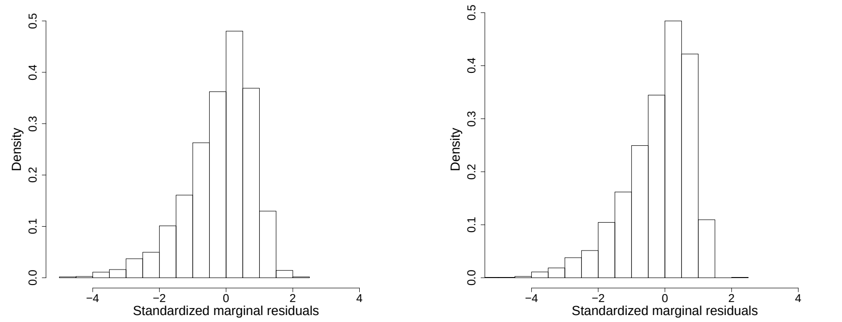


(a) ITM (b) EAM

Figure 9 *Oil well data - Standardized marginal residuals ($\hat{\boldsymbol{\xi}}_i^*$) vs fitted values.*

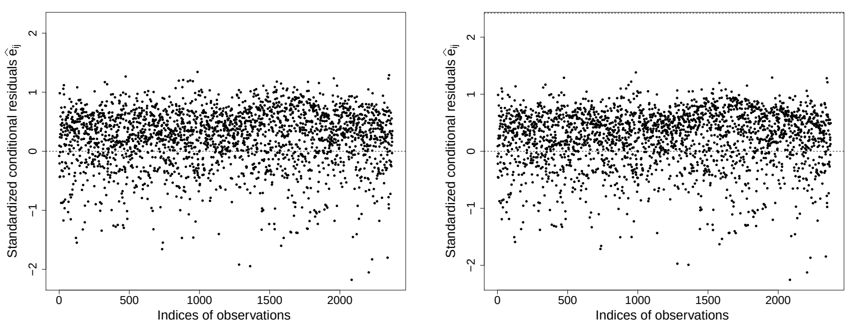
6 Conclusion

There are few proposals for residual analysis in the context of correlated survival data. In particular, we are not aware of any techniques to evaluate validity of the assumptions usually considered in AFT random effects models for the analysis of correlated censored lifetime data. For such purposes, we consider the use of residual plots originally developed for mixed models



(a) ITM (b) EAM

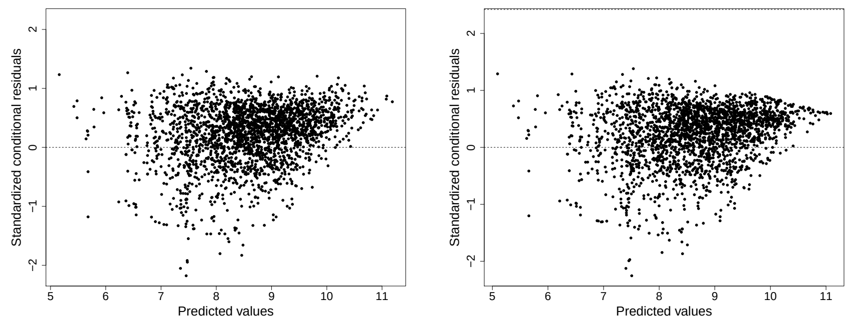
Figure 10 *Oil well data - Histogram for standardized marginal residuals* ($\hat{\boldsymbol{\xi}}_i^*$).



(a) ITM (b) EAM

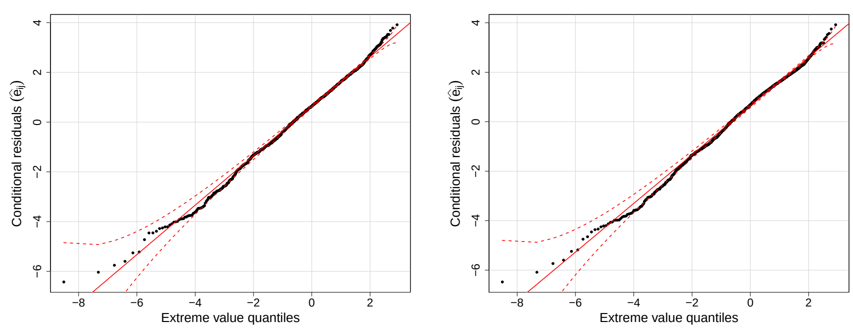
Figure 11 *Oil well data - Standardized conditional residuals* ($\hat{e}_k^*$) *unit index plots.*

to assess the assumptions of AFT models with random effects used to analyze correlated censored lifetime data. Because censoring can compromise the analysis of residuals, we propose and compare two imputation procedures (one based on Ageel (2002): **EAM**; and the other on the method of inverse transform: **ITM**) by means of a simulation study. In the simulation we perceive that **ITM** exceeds **EAM** and is clearly indicated for the imputation of censored data because it frequently avoids the false appearance of violation with respect to the assumptions of linearity, homoskedasticity and correlation structure, that arise by the presence of censorship in the data. However, we also note (through simulations not displayed here for space considerations), that the ITM could have low detection capability with respect



(a) ITM (b) EAM

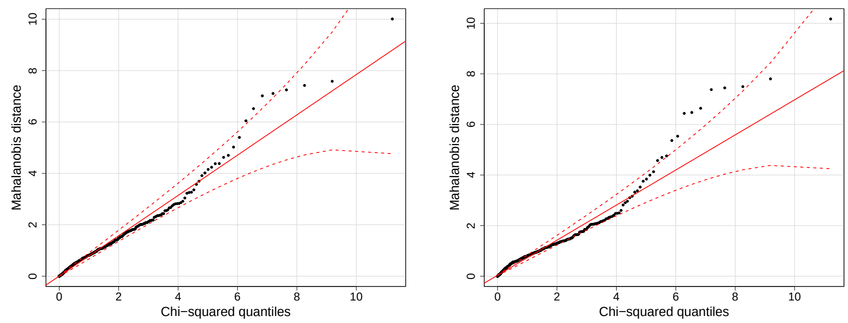
Figure 12 *Oil well data - Standardized conditional residuals ($\hat{e}^*_k$) vs fitted values.*



(a) ITM (b) EAM

Figure 13 *Oil well data - QQ plot for standardized conditional residual ($\hat{e}^*_k$).*

to the distribution of conditional residuals in the context of misspecification, when censorship is high (at least 50 %). This occurs because the censored observations are replaced by values of the assumed model, which makes it difficult to detect an incorrect assumption. An alternative would be use a nonparametric bootstrap procedure to generate the surrogate value of censorship. This is a topic to be investigated in future works. We applied these techniques to evaluate data related to the time between failures of oil wells and observed a similar behaviour of the imputed data by both methods, probably due to the low proportion of censored observations in the data (around 24%).



(a) ITM (b) EAM

Figure 14 *Oil wells data - QQ for Mahalanobis distance ($\mathcal{M}_i$).*

References

Anderson, T.W. (1962) On the distribution of the two-sample Cramer-von Mises criterion. *Annals of Mathematical Statistics*, **33**, 1148–1159

Ageel, M. I. (2002). A novel means of estimating quantiles for 2-parameter weibull distribution under the right random censoring model. *Journal of Computational and Applied Mathematics* **149**, 373–380.

Ascher, H. and Feingold, H. (1984). *Repairable systems reliability: modeling, inference, misconceptions and their causes*. New York: Marcel Dekker.

Bolfarine, H. and Valença, D. M. (2005). Testing homogeneity in weibull-regression models. *Biometrical Journal* **47**, 707–720.

Carvalho, J. B., Valença, D. M. and Singer, J. M. (2014). Prediction of failure probability of oil wells. *Brazilian Journal of Probability and Statistics* **28**, 275–287.

Cramer, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal* **1**, 13–74.

Dobson, A. and Henderson, R. (2003). Diagnostics for joint longitudinal and dropout time modeling. *Biometrics* **59**, 741–751.

Hilden-Minton, J. A. (1995). *Multilevel diagnostics for mixed and hierarchical linear models*. PhD Thesis. University of California, Los Angeles.

Jiang, E. and Verbeke, G. (1998). Asymptotic properties of the empirical blup and blue in mixed linear models. *Statistica Sinica* **8**, 861–885.

Keiding, N., Andersen, P.K. and Klein, J.P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16**, 215–224.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge distribuzione. *Inst. Ital. Attuari, Giorn.* **4**, 83–91.

Lambert, P., Collett, D., Kimber, A. and Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine* **23**, 3177–3192.

Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd Ed. New York: Wiley.

Lawless, J. F. and Thiagarajah, K. (1996). A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics* **38**, 131–138.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **54**, 570–582.

Martinez, E. Z.; Louzada-Neto, F.; Pereira, B. B. (2003). A curva ROC para Testes DiagnÃşsticos. *Cadernos SaÃžde Coletiva* **11**, 7–31.

Nobre, J. S. and Singer, J. M. (2007). Residual analysis for linear mixed models. *Biometrical Journal* **49**, 863–875.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press.

Percy, D. and Alkali, B. (2007). Scheduling preventive maintenance of oil pumps using generalised proportional intensities models. *International Transactions in Operational Research* **14**, 547–563.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects in S and S-PLUS*. New York: Springer.

R Development Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Rao, C. R. (1971a). Estimation of variance and covariance components—MINQUE theory. *Journal of Multivariate Analysis* **1**, 257–275.

Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* **1**, 445–456.

Rizopoulos, D. (2010). Jm: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* **35**, 1–33.

Rizopoulos, D., Verbeke, G. and Lesaffre, E. (2009). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* **66**, 20–29.

Robinson, G. K. (1991). That blup is a good thing: the estimation of random effects. *Statistical Science* **6**, 15–51.

Rubner, Y., Tomasi, C. and Guibas, L.J. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, **40**, 99–121.

Santos, P. B. and Valença, D. M. (2012). Accelerated failure time models with random effects to estimate operation time of oil wells. *Advances and Applications in Statistics* **28**, 23–44.

Searle, S., Casella, G. and Mcculloch, C. E. (1992). *Variance Componentes*. New York: Wiley.

Singer, J.M., Rocha, F.M.M. and Nobre, J.S. (2017). Graphical tools for detecting departures from linear mixed models assumptions and some remedial measures. *International Statistical Review* **85**, 290-324.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, 279–281.

Stewart, W. J. (2009). *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press.

Swain, M.J. and Ballard, D.H. (1991). Color indexing. *International journal of computer vision* **7**, 11–32.

Therneau, T. M., Grambsch, P. M. and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12**, 156–175.

Verbeke, G. and Lesaffre, E. (1996a). Linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association* **91**, 217–221.

Von Mises, R. (1928). *Statistik und wahrheit*. Julius Springer.

Universidade de São Paulo and Universidade Federal do Rio Grande do Norte,
E-mail: elisr@ime.usp.br; dione@ccet.ufrn.br

Universidade de São Paulo,
URL: https://www.ime.usp.br/ jmsinger/doku.p
E-mail: jmsinger@ime.usp.br