Submitted to the Brazilian Journal of Probability and Statistics

Hierarchical modelling of power law processes for the analysis of repairable systems with different truncation times: An empirical Bayes approach

Rodrigo Citton P. dos Reis[a], **Enrico A. Colosimo**[a] **and Gustavo L. Gilardoni**[b]

[a] *Universidade Federal de Minas Gerais*
[b] *Universidade de Brasília*

Abstract. In the data analysis from multiple repairable systems it is usual to observe both different truncation times and heterogeneity among the systems. Among other reasons, the latter is caused by different manufacturing lines and maintenance teams of the systems. In this paper, a hierarchical model is proposed for the statistical analysis of multiple repairable systems under different truncation times. A reparameterization of the power law process is proposed in order to obtain a quasi-conjugate bayesian analysis. An empirical Bayes approach is used to estimate model hyperparameters. The uncertainty in the estimate of these quantities are corrected by using a parametric bootstrap approach. The results are illustrated in a real data set of failure times of power transformers from an electric company in Brazil.

1 Introduction

An issue of interest to statisticians and engineers in the analysis of repairable systems data is how to model the changes in the performance of the system caused by the failure and/or maintenance process. This involves usually a stochastic point process (Andersen et al., 1993; Cook and Lawless, 2007) and statistical analysis (Rigdon and Basu, 2000; Lindqvist, 2006). In the data from multiple repairable systems one observes usually different truncation times and heterogeneity among them. The latter is due to causes such as different locations, manufacturing lines and maintenance teams of the systems, among others. An interesting example of the joint presence of heterogeneity and different truncation times is provided by the power transformers of the electric company of Minas Gerais state in Brazil. These data were first reported and analyzed by Gilardoni and Colosimo (2007). Table 1 contains failure times from forty power transformers, recorded between January 1999

Keywords and phrases. Bootstrap correction, maximum a posterior density, minimal repair, multiple repairable systems, rejection sampling, reliability.

and July 2001. The data consist of the number of failures and failure and truncation times for the forty systems.

Table 1 *Power transformers data.*

System	Number of failures	Failure times (hours)		Trucation times	System	Number of failures	Failure times (hours)		Trucation times
1	2	8,839	17,057	21,887	17	1	15,524		21,886
2	2	9,280	16,442	21,887	18	0			21,440
3	1	10,445		13,533	19	0			369
4	0			7,902	20	2	11,664	17,031	21,857
5	0			8,414	21	0			7,544
6	0			13,331	22	0			6,039
7	1	17,156		21,887	23	1	2,168		6,698
8	1	16,305		21,887	24	1	18,840		21,879
9	1	16,802		21,887	25	0			2,288
10	0			4,881	26	0			2,499
11	0			16,625	27	1	10,668		16,838
12	2	7,396	7,541	19,590	28	1	15,550		21,887
13	0			2,121	29	0			1,616
14	2	15,821	19,746	19,877	30	1	14,041		20,004
15	0			1,927	31 - 40	0			21,888
16	1	15,813		21,886					

Power transformers are complex systems with a large number of components. These devices usually fail because of just one of these components. After this component is repaired, it is expected that the reliability of the transformer does not change. This type of repair is known as *minimal repair*. A failure process that undergoes minimal repair actions is modeled by a nonhomogeneous Poisson process (NHPP) (Baker, 1996). Succinctly, define $N(t)$ to be the number of failures in the interval $(0, t]$. A process $\{N(t) : t \geq 0\}$ having independent increments and starting at $N(0) = 0$ is said to be a Poisson process with intensity $\lambda(\cdot)$ if, for any t, the random variable $N(t)$ follows a Poisson distribution with mean $\Lambda(t) = \int_0^t \lambda(u)du$. The NHPP is a Poisson process with a nonconstant intensity function $\lambda(\cdot)$. In the repairable system literature, the most popular parametric form for λ is the power law process (PLP),

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \tag{1.1}$$

where β and θ are respectively shape and scale parameters. The corresponding mean function is

$$\Lambda(t) = \mathrm{E}\left[N(t)\right] = \int_0^t \lambda(u)\, du = \left(\frac{t}{\theta}\right)^{\beta}. \tag{1.2}$$

The popularity of the PLP model stems from both its mathematical simplicity and its flexibility, in the sense that (1.1) can accommodate situations where the systems either deteriorates ($\beta > 1$) or improves ($\beta < 1$) with time.

When observing data from a single system truncated at τ, the joint likelihood of the number of failures $n = N(\tau)$ and the failure times $0 < t_1 < \cdots < t_n < \tau$ is obtained after noting that $N(\tau)$ follows a Poisson distribution with mean $\Lambda(\tau)$ and, conditional on $N(\tau) = n$, the failure times have the same distribution as the order statistics of a sample of size n from the pdf $g(t) = [\lambda(t)/\Lambda(\tau)]\, I(0 < t < \tau)$, which in the PLP case becomes $g(t) = (\beta/t)(t/\tau)^\beta\, I(0 < t < \tau)$ (see, for instance, Rigdon and Basu, 2000). Therefore,

$$p(n; t_1, \ldots, t_n \,|\, \beta, \theta) = \exp\{-(\tau/\theta)^\beta\}\, \frac{\beta^n}{\theta^{n\beta}} \prod_{j=1}^{n} t_j^{\beta-1}\,. \tag{1.3}$$

(As usual, we assume here and throughout that empty sums and products are equal respectively to zero and one, so that (1.3) becomes $\exp\{-(\tau/\theta)^\beta\}$ when $n = 0$.) If we reparametrize the model in terms of β and $\eta = \mathrm{E}\,[N(\tau)] = (\tau/\theta)^\beta$, the likelihood (1.3) becomes

$$p(n; t_1, \ldots, t_n \,|\, \beta, \eta) \propto \gamma(\eta \,|\, n+1, 1) \times \gamma(\beta \,|\, n+1, w)\,, \tag{1.4}$$

where $w = \sum_{j=1}^{n} \log(\tau/t_j)$ and $\gamma(x \,|\, a, b) = b^a\, x^{a-1} e^{-bx}/\Gamma(a)$ is the density of the gamma distribution with mean a/b and variance a/b^2. The fact that β and η are orthogonal and the striking simplicity of (1.4) makes the (β, η) parameterization quite convenient. It has been used previously by Oliveira, Colosimo and Gilardoni (2012) in nonhierarchical modelling and Ryan, Hamada and Reese (2011) in the context of hierarchical models when all the truncation times are equal. Using either (1.3) or (1.4) it is easy to show then that the maximum likelihood estimates (MLEs) are $\hat\eta = n$ and, provided that $n > 0$, $\hat\beta = n/\sum_{j=1}^{n} \log(\tau/t_j) = n/w$ and $\hat\theta = \tau/n^{1/\hat\beta}$ (the MLEs of β and θ do not exist when $n = 0$). We note that, in the sequel, we will denote (1.3) by writing that $(n; t_1, \ldots, t_n) \sim PLP_\tau(n; t_1, \ldots, t_n \,|\, \beta, \theta)$.

An important aspect to consider regarding the power transformers data in Table 1 is the fact that these systems are located in different places along the Brazilian state of Minas Gerais. Thus, due to climate changes along this state, it is expected that they are exposed to different operating conditions. Therefore, rather than assuming that all 40 systems have the same (β, θ) parameters as in Oliveira, Colosimo and Gilardoni (2012), an individual analysis of each system may be adequate. In other words, one may compute estimates $(\hat\beta_i, \hat\theta_i)$ for each of the 16 systems having $n_i > 0$. Figure 1 shows estimates for the intensity and mean functions (1.1) and (1.2) obtained by substituting the parameters by its MLEs. One can observe that the estimated intensities show quite different behavior (decreasing, concave increasing and convex increasing). While this may be because each system has its unique characteristics, it is more likely the consequence of the fact that the individual estimates are highly inaccurate because the number of observed failures for each system is very small. On the other hand, most of the systems seems to be ageing, but each one in its own way. A hierarchical model which considers this similarity between systems may be more realistic and, at the same time, it would allow to *borrowing* information across systems (Arab, Rigdon and Basu, 2012; Rigdon and Basu, 2000). In other words, the choice by a hierarchical model is a balance between the assumption that the intensity is the same for all power transformers and the one that each transformer has its own intensity.

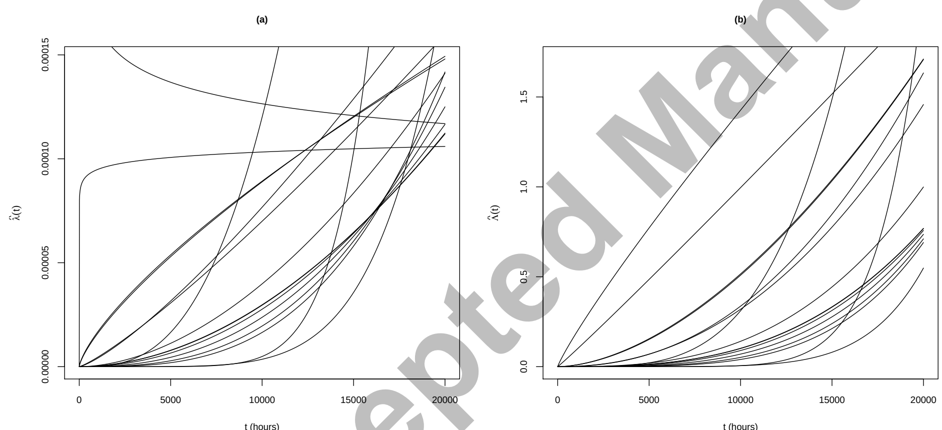

Figure 1 *Maximum likelihood estimates of the intensity* ***(a)*** *and mean* ***(b)*** *functions for the sixteen transformers with* $n_i > 0$.

The objective of this paper is to discuss a hierarchical model to analyze several repairable systems truncated at possible different times. More precisely, the first stage specifies a distribution for the failure times data conditional on the parameters of the PLP, while the second stage specifies a prior distribution for these parameters. Therefore, the specific features of each transformer are modeled in the first stage, while characteristics that are common to all transformers are taken into consideration in the second one. Although there has been some recent interest in the area of hierarchical modeling of repairable systems (see for instance Bhattacharjee, Arjas and Pulkkinen, 2003; Pan and Rigdon, 2009; Ryan, Hamada and Reese, 2011), statistical modeling and inference procedures for the case of multiple repairable systems with different truncation times are still under consideration in the literature. Lindqvist, Elvebakk and Heggland (2003) have considered the issue of unobserved heterogeneity between systems. A counting process, representing the unity, is assumed to be the same for the systems, but their intensity is taken to be different for each one by introducing a frailty term in the model. Frailty affects only the scale parameter of the PLP intensity (see Lawless, 1987). Our model allows both the scale and shape parameters to vary among systems. Following Guida and Pulcini (2005), Giorgio, Guida and Pulcini (2014) used a generalization of the prior proposed by Huang (2001) to model shape and scale parameter of the PLP intensity. The resulting prior depends upon five hyperparameters, one more than our prior model. Furthermore, their approach differs from ours in the sense that they estimate the five hyperparameters using the actual data to elicit an informative prior for a future analysis. It was adopded an empirical Bayes approach to estimate model parameters and hyperparameters. This approach has some advantages in comparison of the fully Bayesian and maximum likehood ones. Empirical Bayes is an approximation to a fully Bayesian approach that provides significant simplifications in computational terms and it allows estimates for the parameters of a system without failures, while there is no maximum likelihood estimate in such case.

The rest of the paper is organized as follows. Section 2 describes the hierarchical model with special focus on the second stage distribution. More precisely, we argue that the (β, η) parameterization together with different truncation times implies that one cannot assume *exchangeability* and suggest a way to overcome this difficulty. Section 3 discusses an empirical Bayes strategy based on maximum posterior density or, equivalently, penalized likelihood estimation for the hyperparameters and, once that the hyperparameters have been estimated, an efficient rejection sampling strategy to obtain *iid* samples from the posterior distribution of the system-specific parameters . Section 3 also presents an implementation of a bootstrap procedure, suggested by Laird and Louis (1987), to correct for the underestimation of uncertainty inherent to the empirical Bayes approach. Section 4 contains an analysis of the power transformers data set, including estimation of the optimal maintenance period under a block maintenance policy. Finally, some conclusions are given in Section 5 and Appendix A describes how to obtain starting values for the penalized likelihood maximization used to estimate the hyperparameters.

2 A hierarchical PLP model

We follow Guida, Calabria and Pulcini (1989), Oliveira, Colosimo and Gilardoni (2012) and Ryan, Hamada and Reese (2011) and parametrize the PLPs in terms of β_i and $\eta_i = \Lambda_i(\tau_i) = (\tau_i/\theta_i)^{\beta_i}$, mainly in view of the simplifications that result from (1.4) and the consequent orthogonality. Of course, it is possible to go from one parameterization to the other provided that one multiplies both prior and posteriors by the appropriate jacobian.

Let $D_i = (n_i; t_{i1}, \ldots, t_{i,n_i})$, $i = 1, \ldots, K$, where K is the number of observed systems, $\boldsymbol{D} = (D_1, \ldots, D_K)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)$. Assuming all throughout conditional independence across systems, the *data level* of the hierarchical model states that

$$p(\boldsymbol{D} \,|\, \boldsymbol{\beta}, \boldsymbol{\eta}) \propto \prod_{i=1}^{K} \gamma(\eta_i \,|\, n_i + 1, 1) \times \gamma(\beta_i \,|\, n_i + 1, w_i)\,, \tag{2.1}$$

where $w_i = \sum_{j=1}^{n_i} \log(\tau_i/t_{ij})$. In other words, data from the i-th system comes from a PLP with parameters β_i and $\theta_i = \tau_i\, \eta_i^{-1/\beta_i}$ observed up to time τ_i [cf. equations (1.3) and (1.4)]. To specify the *prior level* of the model we denote by $\boldsymbol{\phi} = (a_\beta, \beta_0, a_\eta, \theta_0)$ the set of hyperparameters and let

$$p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{\phi}) = \prod_{i=1}^{K} \gamma(\beta_i \,|\, a_\beta, a_\beta/\beta_0) \times \gamma(\eta_i \,|\, a_\eta, a_\eta(\theta_0/\tau_i)^{\beta_i})\,. \tag{2.2}$$

More specifically, we set β_i to follow a gamma distribution with mean β_0 and coefficient of variation $1/\sqrt{a_\beta}$ and, conditional on β_i, η_i follows also a gamma distribution with mean $(\tau_i/\theta_0)^{\beta_i}$ and coefficient of variation $1/\sqrt{a_\eta}$, so that β_0 and θ_0 can be thought off as prior guesses for the β_i's and the θ_i's and a_β and a_η are hyperparameters that control the precision of those prior guesses.

The rationale behind the prior distribution (2.2) can be explained as follows. We begin by noting that it follows from (1.4) that, in the case of a single system, the natural prior for the pair (β, η) is a product of gamma distributions of the form $\gamma(\beta \,|\, a_\beta, a_\beta/\beta_0) \times \gamma(\eta \,|\, a_\eta, a_\eta/\eta_0)$ (cf. Oliveira, Colosimo and Gilardoni, 2012). Following this idea, Ryan, Hamada and Reese (2011) consider a hierarchical model for several PLPs all truncated at the same time $\tau_1 = \ldots = \tau_K = \tau$ and specify the prior level distribution also as a product of gamma distributions of the form $\prod_{i=1}^{K} \gamma(\beta_i \,|\, a_\beta, a_\beta/\beta_0) \times \gamma(\eta_i \,|\, a_\eta, a_\eta/\eta_0)$. However, this possibility does not seem appropriate when the systems have different truncation times, in the sense that it would imply that the pairs

(β_i, η_i) $(i = 1, \ldots, K)$ are *exchangeable*, while one would expect larger values of $\eta_i = \mathrm{E}\,[N_i(\tau_i)]$ for those systems which are observed longer (i.e. which have large τ_i). Although assuming the η_i's to be exchangeable is not reasonable because their definition involves the τ_i's, which are different, it makes sense to assume that the θ_i's are exchangeable irrespective of the truncation times, because their definition (namely, θ_i is the time such that $\mathrm{E}\,[N_i(\theta_i)] = 1$) does not involve the τ_i's. Therefore, we want the prior level distribution $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{\phi})$ to be such that the pairs $(\beta_i, \theta_i = \tau_i\, \eta_i^{-1/\beta_i})$ are exchangeable. Now, it is straightforward to check that (2.2) implies that

$$p(\boldsymbol{\beta}, \boldsymbol{\theta} \,|\, \boldsymbol{\phi}) = \prod_{i=1}^{K} \gamma(\beta_i \,|\, a_\beta, a_\beta/\beta_0) \times \frac{a_\eta^{a_\eta}}{\Gamma(a_\eta)} \frac{\beta_i}{\theta_i} \left(\frac{\theta_0}{\theta_i}\right)^{a_\eta \beta_i} \exp\{-a_\eta(\theta_0/\theta_i)^{\beta_i}\}\,,$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$. Since the truncation times τ_i do not appear in the right hand side of this last expression, this implies that the pairs (β_i, θ_i) are indeed exchangeable.

An alternative derivation of (2.2) is as follows. Write $p(\beta_i, \eta_i | \boldsymbol{\phi}) = p(\beta_i | \boldsymbol{\phi}) \times p(\eta_i | \beta_i, \boldsymbol{\phi})$ and suppose that one wants to set $\beta_i \,|\, \boldsymbol{\phi} \sim \mathrm{Gamma}(a_\beta, a_\beta/\beta_0)$ and $\eta_i \,|\, \beta_i, \boldsymbol{\phi} \sim \mathrm{Gamma}(a_\eta, b_\eta)$, where a_η and b_η could possibly depend on β_i and τ_i. Then the β_i's are exchangeable and a necessary condition for the pairs (β_i, θ_i) to be exchangeable is that $\mathrm{E}\,[\theta_i^{-\beta_i} | \boldsymbol{\phi}]$ does not depend on the system i. Now, since $\theta_i^{-\beta_i} = \tau_i^{-\beta_i}\, \eta_i$,

$$\mathrm{E}\,[\theta_i^{-\beta_i} \,|\boldsymbol{\phi}] = \mathrm{E}\,\left[\mathrm{E}\,[\tau_i^{-\beta_i}\, \eta_i \,|\, \beta_i, \boldsymbol{\phi}]\right] = \mathrm{E}\,[\tau_i^{-\beta_i}\,(a_\eta/b_\eta) \,|\, \boldsymbol{\phi}]\,.$$

It is easy to see that for this not to depend on τ_i, it is necessary that there exists a function h such that $\mathrm{E}\,[\tau_i^{-\beta_i}\,(a_\eta/b_\eta) \,|\, \boldsymbol{\phi}] = h(\beta_i)$. The prior $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{\phi})$ given in (2.2) corresponds to the choice $h(\beta_i) = \theta_0^{-\beta_i}$. In other words, the previous argument shows that for the prior (2.2) one has that $\mathrm{E}\,[\theta_i^{-\beta_i} \,|\boldsymbol{\phi}] = \mathrm{E}\,[\theta_0^{-\beta_i} |\boldsymbol{\phi}]$, showing again why θ_0 can be thought of as a prior guess for the θ_i's.

To complete the specification of the hierarchical model, we assume an independent prior distribution for the hyperparameters of the form

$$p(\boldsymbol{\phi}) = p(a_\beta) \times p(\beta_0) \times p(a_\eta) \times p(\theta_0) \propto \exp\{-\xi_1 a_\beta\}\, \exp\{-\xi_2 a_\eta\}\,, \quad (2.3)$$

i.e., we set both $p(\beta_0) \propto 1$ and $p(\theta_0) \propto 1$ and exponential densities with means ξ_1^{-1} and ξ_2^{-1} respectively for a_β and a_η. The exponential distribution is a common choice for the shape parameter of the Gamma-Poisson hierarchical model (see for example George, Makov and Smith (1993), and related applications Pérez, Martín and Rufo (2006); Pesaran, Pettenuzzo and Timmermann (2006); Perkins et al. (2012)), that can be thought as a prototype for the PLP hierarchical model. In Section 3 we discuss the specification of ξ_1 and ξ_2.

In the rest of the paper we discuss an empirical Bayes procedure which estimates $\boldsymbol{\phi}$ from data by maximizing the posterior density $p(\boldsymbol{\phi}|\boldsymbol{D})$ or, equivalently, by maximizing a penalized likelihood (see Section 3 and Appendix A). Once that an estimate $\hat{\boldsymbol{\phi}}$ has been obtained, inferences about quantities specific to each system proceeds straightforward after noting from (2.1) and (2.2) that

$$p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi}) = \prod_{i=1}^{K} p(\eta_i \,|\, \beta_i, D_i, \boldsymbol{\phi}) \times p(\beta_i \,|\, D_i, \boldsymbol{\phi}), \tag{2.4}$$

where

$$p(\eta_i \,|\, \beta_i, D_i, \boldsymbol{\phi}) = \gamma(\eta_i \,|\, a_\eta + n_i, a_\eta\,(\theta_0/\tau_i)^{\beta_i} + 1), \tag{2.5}$$

and

$$p(\beta_i \,|\, D_i, \boldsymbol{\phi}) \propto \gamma(\beta_i \,|\, a_\beta + n_i, a_\beta/\beta_0 + w_i) \times \frac{[a_\eta(\theta_0/\tau_i)^{\beta_i}]^{a_\eta}}{[a_\eta(\theta_0/\tau_i)^{\beta_i} + 1]^{a_\eta + n_i}}. \tag{2.6}$$

3 Empirical Bayes inference for the hierarchical PLP model

To make inferences for the hierarchical PLP model we adopt a parametric empirical Bayes (PEB) approach. The PEB approach uses the observed data to estimate, usually by the maximum likelihood method, the hyperparameters $\boldsymbol{\phi} = (a_\beta, \beta_0, a_\eta, \theta_0)$. Then, one replaces $\boldsymbol{\phi}$ by its estimate $\hat{\boldsymbol{\phi}}$ in the conditional posterior (2.4)–(2.6) to make inferences with respect to $(\boldsymbol{\beta}, \boldsymbol{\eta})$. However, this appproach ignores the uncertainty in the estimation of $\boldsymbol{\phi}$, and hence tends to underestimate variances and produce too narrow intervals (Carlin and Gelfand, 1990). Kass and Steffey (1989) proposed first and second order approximations to Var $[h(\beta_i, \eta_i)|\mathbf{D}]$ which requires computation of higher order derivatives of the marginal log-likelihood. Computation of these derivatives becomes too complex for the hierarchical PLP case. Hence, we followed the proposal of Laird and Louis (1987) and use a parametric bootstrap to approximate the marginal posterior distribution of $\boldsymbol{\phi}$. For details about the PEB approach see, for instance, Morris (1983), Casella (1985) or, in the reliability literature, Gaver and O'Muircheartaigh (1987).

This section is divided into three subsections which discuss respectively (i) the maximum posterior density estimate for $\boldsymbol{\phi}$, (ii) a rejection sampling algorithm to sample from the conditional posterior $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$ and (iii) the parametric bootstrap strategy used to approximate the posterior marginal distribution $p(\boldsymbol{\phi}|\boldsymbol{D})$ which is then used to correct both standard errors of point estimates and credibility intervals for the system specific parameters.

3.1 Maximum posterior density estimate

From (2.1) and (2.2), the marginal likelihood for $\boldsymbol{\phi}$ is given by

$$
\begin{aligned}
p(\mathbf{D}|\boldsymbol{\phi}) &= \int_{\mathbb{R}_+^K}\int_{\mathbb{R}_+^K} p(\mathbf{D}|\boldsymbol{\beta},\boldsymbol{\eta}) \times p(\boldsymbol{\beta},\boldsymbol{\eta}|\boldsymbol{\phi}) d\boldsymbol{\eta} d\boldsymbol{\beta} \\
&= \prod_{i=1}^{K} \left(\prod_{j=1}^{n_i} \frac{1}{t_{ij}} \right) \frac{\Gamma(a_\eta + n_i)}{\Gamma(a_\eta)\Gamma(a_\beta)} \left(\frac{a_\beta}{\beta_0}\right)^{a_\beta} \\
&\quad \times \int_0^\infty \left[\frac{a_\eta(\theta_0/\tau_i)^{\beta_i}}{a_\eta(\theta_0/\tau_i)^{\beta_i}+1} \right]^{a_\eta} \left[\frac{1}{a_\eta(\theta_0/\tau_i)^{\beta_i}+1} \right]^{n_i} \\
&\qquad \times \beta_i^{a_\beta+n_i-1} e^{-\beta_i(a_\beta/\beta_0 + w_i)} d\beta_i. \qquad (3.1)
\end{aligned}
$$

Note that the last integral in (3.1) has no closed form and it should have to be computed numerically in the maximization algorithm. Hence, the marginal posterior distribution of $\boldsymbol{\phi}$ is

$$p(\boldsymbol{\phi}|\mathbf{D}) \propto p(\mathbf{D}|\boldsymbol{\phi}) \times p(\boldsymbol{\phi}), \qquad (3.2)$$

where $p(\boldsymbol{\phi})$ is given in (2.3). Note that maximizing (3.2) is equivalent to maximizing

$$\ell(\boldsymbol{\phi}) = \log p(\mathbf{D}|\boldsymbol{\phi}) - (\xi_1 a_\beta + \xi_2 a_\eta), \qquad (3.3)$$

showing that one could think of the maximum posterior estimate of $\boldsymbol{\phi}$ as a penalized likelihood approach. Maximization of (3.3) is carried out numerically. Initial values to start the algorithm are discussed in Appendix A.

In order to evaluate the behavior of the estimators obtained from the maximization of (3.3), we conducted a Monte Carlo simulation study. The Monte Carlo scenarios were designed to generate data similar to the transformers example. Hence, we set the hyperparameters $\beta_0 = 2$, $\theta_0 = 10,000$, $a_\beta = 2, 10$, $a_\eta = 2, 10$, truncation times varying from 2,000 to 20,000 hours and $K = 10, 40, 70$ and 100 systems. We compared the mean and standard errors of the estimates $(\hat{a}_\beta, \hat{\beta}_0, \hat{a}_\eta, \hat{\theta}_0)$ of 500 Monte Carlo replicates using (i) maximization of the marginal likelihood, (ii) maximization of the marginal posterior of $\boldsymbol{\phi}$ with $\xi_1 = \xi_2 = 1$ and (iii) same as (ii) but with $\xi_1 = \xi_2 = 0.1$. All the results were obtained using the software R, version 3.0.1 (R Core Team, 2013).

The results are summarized in Figures 2–5. Briefly, the estimates for β_0 and θ_0 behave similar for the three methods. In other words, the introduction of a penalty of the form $\xi_1 a_\beta + \xi_2 a_\eta$ does not impact much the estimates of β_0 and θ_0. On the other hand, the estimates of a_β and a_η obtained maximizing the marginal posterior performed better than the ones obtained by maximizing the marginal likelihood, in the sense that they have smaller bias and standard errors for small K. Of the two options $\xi_1 = \xi_2 = 1$ and $\xi_1 = \xi_2 = 0.1$, the latter seems to be slightly better. In terms of the prior distribution (2.3) for $\boldsymbol{\phi}$, this amounts to setting (improper) uniform priors for both β_0 and θ_0 and exponential distributions with mean and standard deviation $1/0.1 = 10$ for both a_β and a_η. We finally note that, as expected, as the amount of information grows (i.e., K grows), the three estimators seem to converge to the true values of $\boldsymbol{\phi}$.

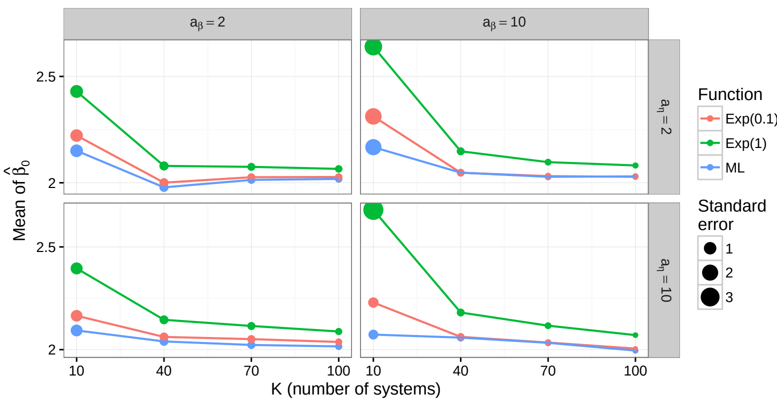

Figure 2 *Mean value of the estimates of* β_0*. Point sizes are proportional to the standard error of the estimates.*

3.2 Simulations for the coditional posterior distribution

For given $\boldsymbol{\phi}$ (e.g. $\hat{\boldsymbol{\phi}}$ obtained by maximizing (3.3)), *iid* simulation from the conditional posterior distribution (2.4)–(2.6) is straightforward using the rejection sampling algorithm (see, for instance, Devroye, 1986; Gelman et al., 2003). Note first that (i) the pairs (β_i, η_i) are conditionally independent and (ii) given β_i, η_i follows a Gamma distribution. Hence, the only difficulty in order to sample from $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$ is how to sample from (2.6).

Let $F(\beta_i)$ be the last factor in the right hand side of (2.6), i.e.

$$F(\beta_i) = \frac{[a_\eta(\theta_0/\tau_i)^{\beta_i}]^{a_\eta}}{[a_\eta(\theta_0/\tau_i)^{\beta_i} + 1]^{a_\eta + n_i}}.$$

Simple algebra shows that $F(\beta_i)$ is maximized when $\beta_i = \beta_i^* = \max\{0, -\log n_i / \log(\theta_0/\tau_i)\}$. Therefore, we can generate a random variable having the pdf (2.6) by

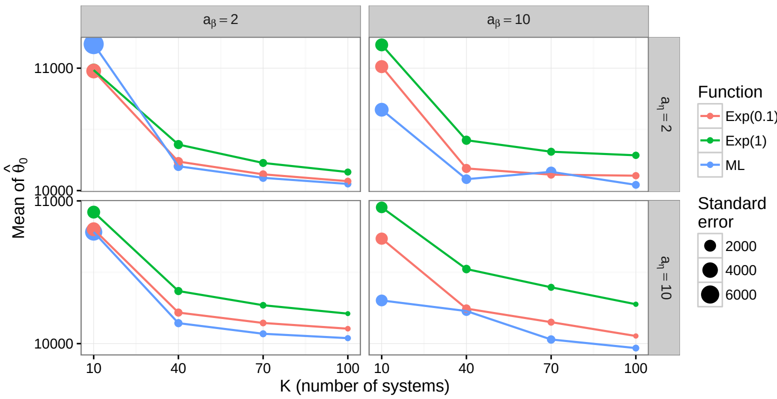

Figure 3 *Mean value of the estimates of θ_0. Point sizes are proportional to the standard error of the estimates.*

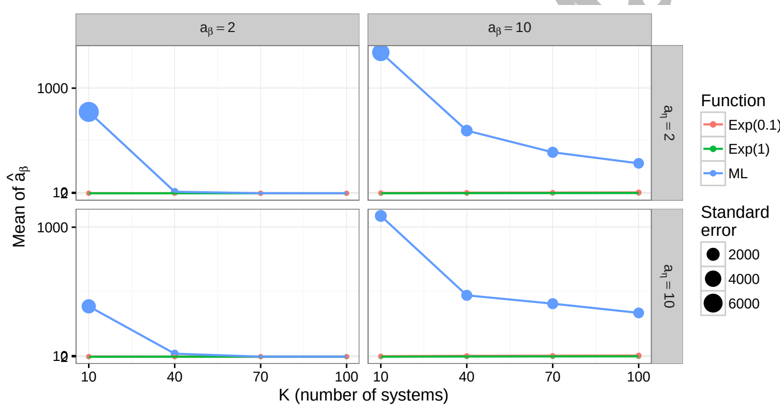

Figure 4 *Mean value of the estimates of a_β. Point sizes are proportional to the standard error of the estimates.*

1. Generate $\beta_i^{(\text{cand})} \sim$ Gamma $(\beta_i | a_\beta + n_i, a_\beta/\beta_0 + w_i)$ and $u \sim$ Uniform(0,1).
2. Define $C_i = F(\beta_i^*)$. If $u\,C_i \le F(\beta_i^{(\text{cand})})$, accept $\beta_i = \beta_i^{(\text{cand})}$. Otherwise, repeat step 1 until the acceptance condition is met.

Using the structure of the model we can then generate an observation from $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$ by running the previous algorithm K times to obtain $\beta_1, \ldots, \beta_K$ and then sampling $\eta_1, \ldots, \eta_K$ from the Gamma distributions (2.5). We then repeat this procedure M times to obtain an *iid* sample $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\eta}^{(1)}), \ldots, (\boldsymbol{\beta}^{(M)}, \boldsymbol{\eta}^{(M)})$ from $p(\boldsymbol{\beta}, \boldsymbol{\eta} \,|\, \boldsymbol{D}, \boldsymbol{\phi})$.

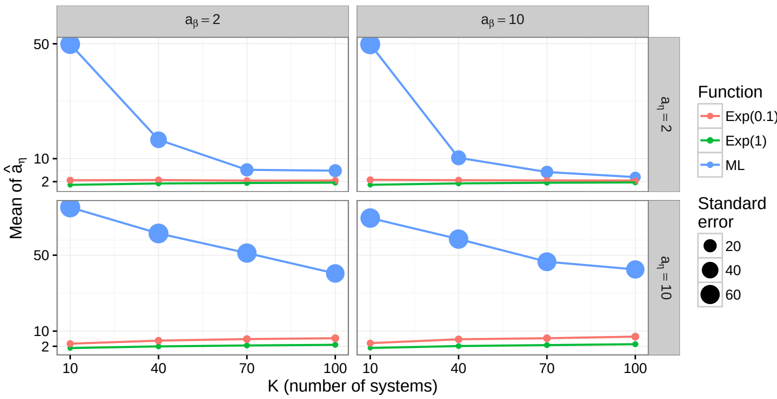

Figure 5 *Mean value of the estimates of a_η. Point sizes are proportional to the standard error of the estimates.*

3.3 Parametric Bootstrap correction

From a Bayesian point of view, the PEB distribution $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\boldsymbol{\phi}})$ is an approximation to the marginal posterior distribution

$$p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = \int_{\mathbb{R}^4_+} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi}) p(\boldsymbol{\phi}|\mathbf{D}) d\boldsymbol{\phi}\,, \tag{3.4}$$

where $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi})$ is given by (2.4) and $p(\boldsymbol{\phi}|\mathbf{D})$ by (3.1)–(3.2). In other words, the PEB approach replaces $p(\boldsymbol{\phi}|\mathbf{D})$ by the Dirac measure (see Schilling, 2005, pg. 26) $\delta_{\hat{\boldsymbol{\phi}}}$ to get

$$\tilde{p}_{\text{naive}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = \int_{\mathbb{R}^4_+} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi}) \delta_{\hat{\boldsymbol{\phi}}}(d\boldsymbol{\phi}) = p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\boldsymbol{\phi}}), \tag{3.5}$$

where $\hat{\boldsymbol{\phi}}$ is the maximum posterior density estimate of $\boldsymbol{\phi}$. This approximation is naive since it fails to take into account the uncertainty with respect to the estimation of $\boldsymbol{\phi}$. Consequently, posterior variances tend to be underestimated and credible intervals too narrow. Laird and Louis (1987) suggested that a more satisfactory solution would be to replace the posterior $p(\boldsymbol{\phi}|\mathbf{D})$ in (3.4) by the sampling distribution $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$ of $\hat{\boldsymbol{\phi}}$. When $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$ is not known or difficult to obtain, they propose to use a parametric bootstrap method to get a proxy for $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$. The bootstrap algorithm obtains bootstrap replications $\hat{\boldsymbol{\phi}}^{(b)}$ $(b = 1 \ldots, B)$ on which to base the approximation to $f_{\hat{\boldsymbol{\phi}}}(\boldsymbol{\phi})$. Given $\hat{\boldsymbol{\phi}}$, the maximum posterior density estimate of $\boldsymbol{\phi}$ using the original data, we generate first $(\boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}^{(b)})$ from the prior distribution $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\hat{\boldsymbol{\phi}})$ and then $\boldsymbol{D}^{(b)}$

from $p(\boldsymbol{D}|\boldsymbol{\beta}^{(b)}, \boldsymbol{\eta}^{(b)})$. Let $\hat{\boldsymbol{\phi}}^{(b)}$ be the maximum posterior density estimate of $\boldsymbol{\phi}$ using the simulated data $\mathbf{D}^{(b)}$, and $\hat{f}_B(\boldsymbol{\phi})$ be the discrete probability function that puts mass $1/B$ on $\hat{\boldsymbol{\phi}}^{(b)}$. The bootstrap corrected approximation to $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ is

$$\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}) = \int_{\mathbb{R}^4_+} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \boldsymbol{\phi})\hat{f}_B(\boldsymbol{\phi})d\boldsymbol{\phi} = \frac{1}{B}\sum_{b=1}^{B} p(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D}, \hat{\boldsymbol{\phi}}^{(b)}). \quad (3.6)$$

An *iid* sample from the bootstrap corrected distribution $\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ is obtained by (i) drawing at random one of the bootstrap replications $\hat{\boldsymbol{\phi}}^{(b)}$ $(b = 1\ \ldots, B)$ and (ii) generate a pair $(\boldsymbol{\beta}, \boldsymbol{\eta})$ from the conditional posterior $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\boldsymbol{D}, \hat{\boldsymbol{\phi}}^{(b)})$ using the drawed value of $\hat{\boldsymbol{\phi}}^{(b)}$ and the algorithm described in Section 3.2.

4 Application: Power transformers data set

We return now to the power transformers data in Table 1. Interest centers in estimation of some quantities associated to the reliability of each system. Among these we mention the β_i's, specifically to assess whether the systems are degrading ($\beta_i > 1$) or improving ($\beta_i < 1$), the scale parameters $\theta_i = \tau_i/\eta_i^{1/\beta_i}$, the probability that no failure occur in a period of time of length l_0 starting at s, called the *reliability function* of the system (Hamada et al., 2008),

$$R_i(s, l_0) = \Pr(N_i(s, s + l_0) = 0|\beta_i, \theta_i) = \exp\left\{\left(\frac{s}{\theta_i}\right)^{\beta_i} - \left(\frac{s + l_0}{\theta_i}\right)^{\beta_i}\right\},$$

where $N_i(s, l_0)$ is the number of failures in the interval (s, l_0) for the i-*th* sytem, for given values of s and l_0 (e.g. $l_0 = 4,380$ and $8,760$ hours, corresponding respectively to 6 months and one year), and, finally, the *optimal maintenance checkpoint* $t^{*(i)}_{PM}$ under a block policy (cf. Mazzuchi and Soyer, 1996), which we explain below.

4.1 Preventive maintenance policy

The optimal maintenance checkpoint relates to the decision of whether to perform a *perfect* preventive maintenance on the system. A perfect preventive maintenance leaves the system in *as good as new* condition and, hence, can also be thought of as the action of replacing the system by a new one. One of the most common strategies of planned preventive maintenance is the block policy. This strategy consists in performing a preventive maintenance at the end of each time interval of length t_{PM}, regardless of the number of previous failures. Under the block policy, the cost per unit of time of the i-th system is

$$C_i(t_{PM}, N_i(t_{PM})) = \frac{C_{PM} + C_{MR} N_i(t_{PM})}{t_{PM}},$$

where $N_i(t_{PM})$ is the number of failures of the i-th system in the time interval of length t_{PM}, C_{PM} is the cost of the preventive maintenance, and C_{MR} is the cost of a minimal repair (unscheduled maintenance due to a failure). Since $N_i(t_{PM})$ is a random quantity, we obtain the conditional expected cost per time unit given (β_i, η_i) as

$$\mathrm{E}\left[C_i(t_{PM}, N_i(t_{PM}))|\beta_i, \eta_i\right] = \frac{C_{PM} + C_{MR}\Lambda_i(t_{PM})}{t_{PM}}. \tag{4.1}$$

A classical approach takes the optimal maintenance time to be the time that minimize (4.1) and compute an estimate replacing (β_i, η_i) by their estimates (see, for instance, Barlow and Hunter, 1960; Gilardoni and Colosimo, 2007, 2011; Oliveira, Colosimo and Gilardoni, 2012; Gilardoni, Oliveira and Colosimo, 2013). Here, instead, we follow Mazzuchi and Soyer (1996) taking the optimal maintenance time $t_{PM}^{*(i)}$ as the value t_{PM} that minimizes the expected cost

$$\mathrm{E}\left[C_i(t_{PM}, N_i(t_{PM}))\right] = \int \frac{C_{PM} + C_{MR}\eta_i(t_{PM}/\tau_i)^{\beta_i}}{t_{PM}} p(\beta_i, \eta_i|D_i) d\beta_i d\eta_i. \tag{4.2}$$

In order to compute an estimate of $t_{PM}^{*(i)}$ we use a sample $\{(\beta_i^{(m)}, \eta_i^{(m)})$ $, m = 1, \ldots, M\}$ from the approximate posterior, either $\tilde{p}_{\text{naive}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ or $\tilde{p}_{\text{boot}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$, given in equations (3.5)–(3.6), and approximate the right hand side of (4.2) by $M^{-1}\sum_{m=1}^{M}[C_{PM} + C_{MR}\eta_i^{(m)}(t_{PM}/\tau_i)^{\beta_i^{(m)}}]/t_{PM}$. The estimate of the optimal maintenance checkpoint is then obtained by a numerical minimization procedure.

4.2 Results

The maximum posterior density estimates of the hyperparameters were obtained maximizing Equation (3.3) with $\xi_1 = \xi_2 = 0.1$. This gave $\hat{\boldsymbol{\phi}} = (\hat{a}_\beta, \hat{\beta}_0, \hat{a}_\eta, \hat{\theta}_0) = (7.02; 2.29; 4.71; 23,980)$. Using this estimates we then generated a sample of size $M = 10,000$ from both $\tilde{p}_{\text{naive}}(\boldsymbol{\beta}, \boldsymbol{\eta}|\mathbf{D})$ and $\tilde{p}_{\text{boot}}(\boldsymbol{\beta}$ $, \boldsymbol{\eta}|\mathbf{D})$, where for the latter it was used $B = 1,000$. Approximations to the estimates of the quantities of interests under squared error loss were then computed by taking the posterior sample averages of the corresponding functions. Likewise, approximate *high posterior density* (HPD) intervals were computed taking the sampling quantiles, say a and $(1-b)$, so that $(1-a-b)$ gives the desired coverage (posterior probability) and the length of the interval is minimum.

Table 2 shows the maximum likelihood and PEB estimates of the β_i and η_i. Note that, unlike the ML approach, in the hierarchical approach estimates of β_i are obtained even for the systems that have no failures. Furthermore, note that the PEB estimates of β_i are a compromise between the ML estimates, which use only data from the i-th system, and the estimated prior mean of β_i, $\hat{\beta}_0$, which uses data from all systems. For the systems with $n_i = 0$, $\hat{\beta}_i$ is close to $\hat{\beta}_0$, since the individual likelihood has little or no information about β_i.

Table 2 *Maximum likelihood (MLE), naive and bootstrap PEB estimates of (β_i, η_i) for the power transformers data.*

	β_i					η_i				
		Naive		Bootstrap			Naive		Bootstrap	
System i	MLE	Mean	SD	Mean	SD	MLE	Mean	SD	Mean	SD
1	1.73	2.08	0.69	2.14	0.92	2	1.00	0.40	1.14	0.60
2	1.75	2.09	0.70	2.16	0.94	2	1.01	0.39	1.14	0.59
3	3.86	2.16	0.75	2.29	1.07	1	0.36	0.21	0.40	0.30
4	-	2.35	0.87	2.80	1.85	0	0.10	0.10	0.10	0.13
5	-	2.36	0.87	2.84	1.96	0	0.11	0.11	0.11	0.14
6	-	2.41	0.88	2.88	1.92	0	0.26	0.17	0.23	0.22
7	4.11	2.41	0.84	2.77	1.45	1	0.83	0.35	0.86	0.50
8	3.40	2.39	0.85	2.69	1.41	1	0.83	0.36	0.86	0.50
9	3.78	2.41	0.86	2.74	1.46	1	0.84	0.36	0.86	0.49
10	-	2.33	0.85	2.77	1.84	0	0.05	0.07	0.05	0.09
11	-	2.39	0.87	2.91	2.12	0	0.40	0.22	0.35	0.28
12	1.04	1.73	0.58	1.65	0.68	2	0.88	0.35	1.03	0.55
13	-	2.31	0.87	2.74	1.85	0	0.02	0.03	0.02	0.06
14	8.52	2.55	0.85	2.99	1.43	2	0.79	0.32	0.89	0.51
15	-	2.30	0.87	2.73	1.87	0	0.01	0.03	0.02	0.06
16	3.08	2.35	0.83	2.64	1.35	1	0.83	0.35	0.86	0.50
17	2.91	2.36	0.82	2.63	1.36	1	0.84	0.36	0.87	0.50
18	-	2.34	0.88	2.82	1.97	0	0.66	0.31	0.59	0.41
19	-	2.28	0.86	2.74	1.90	0	0.00	0.01	0.00	0.03
20	2.28	2.24	0.75	2.38	1.02	2	0.99	0.39	1.12	0.60
21	-	2.36	0.87	2.79	1.87	0	0.09	0.09	0.09	0.13
22	-	2.33	0.84	2.78	1.82	0	0.07	0.08	0.07	0.11
23	0.89	1.52	0.52	1.41	0.60	1	0.20	0.15	0.27	0.24
24	6.69	2.49	0.87	2.97	1.69	1	0.83	0.35	0.86	0.50
25	-	2.31	0.85	2.72	1.83	0	0.02	0.03	0.02	0.06
26	-	2.33	0.87	2.74	1.84	0	0.02	0.04	0.03	0.06
27	2.19	2.17	0.76	2.30	1.09	1	0.53	0.26	0.56	0.36
28	2.93	2.35	0.82	2.61	1.32	1	0.83	0.36	0.86	0.50
29	-	2.31	0.86	2.76	2.03	0	0.01	0.02	0.02	0.05
30	2.83	2.31	0.82	2.56	1.28	1	0.71	0.31	0.74	0.44
31 - 40	-	2.34	0.88	2.79	2.03	0	0.69	0.32	0.62	0.42

Table 3 presents PEB estimates for the quantities $\Pr(\beta_i > 1|\hat{\boldsymbol{\phi}})$ and $t^{*(i)}_{PM}$. If we look at the probability that a system is degrading, namely $\tilde{\Pr}(\beta_i > 1|D_i, \hat{\boldsymbol{\phi}})$, the smallest values are 0.742 and 0.845, respectively for systems 23 and 12, while all others are greater than 0.93, indicating strong evidence in the sense that the intensities are increasing and the transformers are degrading with time. This can be seen also in Figure 6, which shows the posterior means of the reliability function for the forty systems. Figure 6(a) shows, for instance, that a system that was followed-up for six months has probability of having no failure in the next six months varying from 0.832 to 0.942. Similarly, Figure 6(b) shows that if a system was followed-up to one year, the probability of observing no failures in the next year vary from 0.604 to 0.783. Note the distinct behavior of the reliability functions of systems 12 and 23. These two systems are the power transformers that presented the earliest failure times. The columns $t^{*(i)}_{PM}$ of Table 3 also show the optimal maintenance check points for each system. To compute this we followed Gilardoni and Colosimo (2007) and Oliveira, Colosimo and Gilardoni (2012), which consider that the cost of a minimal repair is fifteen times the cost of a preventive maintenance. The estimated optimal maintenance checkpoints vary from 6,592 (system 20) to 9,348 hours (system 23). Using the same data, but considering that the forty power transformers are a sample of the same power law process (i.e. same β and θ for all systems), Gilardoni and Colosimo (2007) and Oliveira, Colosimo and Gilardoni (2012), using respectively ML and a Bayesian approach, arrived at an optimal time of about $6,420$ hours. The hierarchical approach has the advantage that each power transformer can be subject to its own optimal maintenance checkpoint, allowing therefore a greater flexibility in the maintenance policy.

Table 3 *PEB estimates for probability that a system is degrading* ($\tilde{\Pr}(\beta_i > 1|D_i, \hat{\boldsymbol{\phi}})$) *and optimal maintenance checkpoints* ($t^{*(i)}_{PM}$) *for the power transformers data.*

System	$\tilde{\Pr}(\beta_i > 1\|D_i, \hat{\boldsymbol{\phi}})$	$t^{*(i)}_{PM}$	System	$\tilde{\Pr}(\beta_i > 1\|D_i, \hat{\boldsymbol{\phi}})$	$t^{*(i)}_{PM}$
1	0.930	6,687	17	0.953	7,642
2	0.933	6,686	18	0.932	9,202
3	0.930	7,019	19	0.931	8,224
4	0.941	8,218	20	0.957	6,592
5	0.947	8,233	21	0.944	8,165
6	0.942	8,508	22	0.942	8,124
7	0.960	7,689	23	0.742	9,348
8	0.952	7,755	24	0.965	7,825
9	0.958	7,743	25	0.933	8,141
10	0.942	8,133	26	0.933	8,138
11	0.938	8,804	27	0.931	7,303
12	0.845	7,291	28	0.955	7,695
13	0.936	8,148	29	0.935	8,168
14	0.981	6,795	30	0.951	7,489
15	0.933	8,181	31-40	0.931	9,295
16	0.956	7,678			

An insight of the bootstrap correction can be seen from the histograms of the bootstrap sample of $\hat{\boldsymbol{\phi}}$ (Figure 7). Note that the sampling distribution of the estimates of the shape parameters a_β and a_η appear to be much more dispersed than those of β_0 and θ_0. The effect of the bootstrap correction can also be seen in Figure 8, which shows the HPD intervals for the β_i and θ_i computed using both the naive and the bootstrap corrected posterior. As expected, the bootstrap correction accounts for wider HPD intervals, which

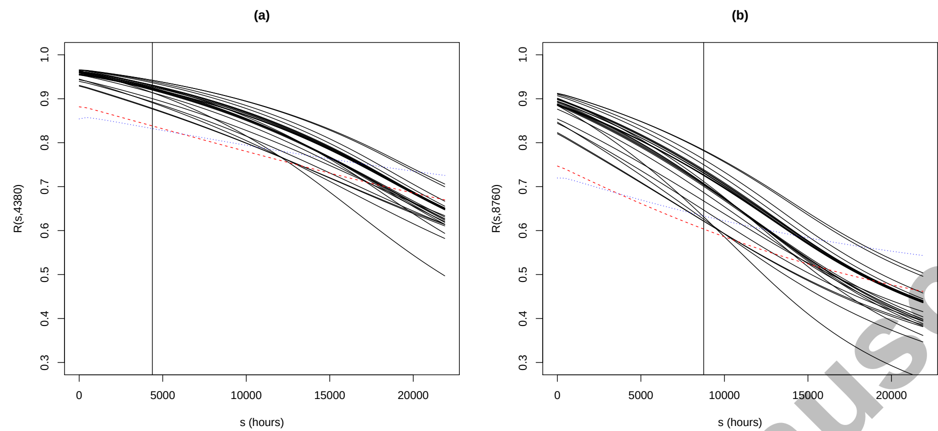

Figure 6 *Posterior means of the reliability function of the forty power transformers when* $l_0 = 4,380$ *hours (6 months)* ***(a)*** *and* $l_0 = 8,760$ *hours (one year)* ***(b)****. The dashed red line and dotted blue line represent respectively systems 12 and 23. Vertical lines represtent* $s = 4,380$ *hours* ***(a)****, and* $s = 8,760$ *hours* ***(b)****.*

we believe reflects better the uncertainty in the data.

To evaluate the impact of the choice of parameters ξ_1 and ξ_2 on the estimates of parameters β_i and θ_i, we performed a sensitivity analysis. Changing the value of the parameters ξ_1 and ξ_2 to 0.5 and 0.02 did not impact on the estimates of parameters of the PLP (Figures 9 and 10). We also considered independent gamma distributions for hyperparameters β_0 and θ_0, with prior mean equal to the starting values of β_0 and θ_0 and different values of prior variance (10, 100 and 1000), instead of uniform (improper) priors. Results are similars for greater variance values.

Finally, in order to understand the behavior of our model, Figure 11 shows the posterior means $\tilde{\beta}_i$ for the parameter β_i, as a function of the prior standard deviation. As the standard deviation of β_i increases, the posterior mean of each β_i moves away in the direction of the ML estimate. On the other hand, as the standard deviation of β_i decreases to zero, the posterior mean of the β_i tend to the common value $\hat{\beta}_0$.

5 Conclusions

A hierarchical model was proposed for the analysis of multiple repairable systems with different truncation times. Scale and shape parameter of the power law intensity function of a nonhomogeneous Poisson proccess are allowed to vary among the systems. A suitable reparameterization was used to obtain

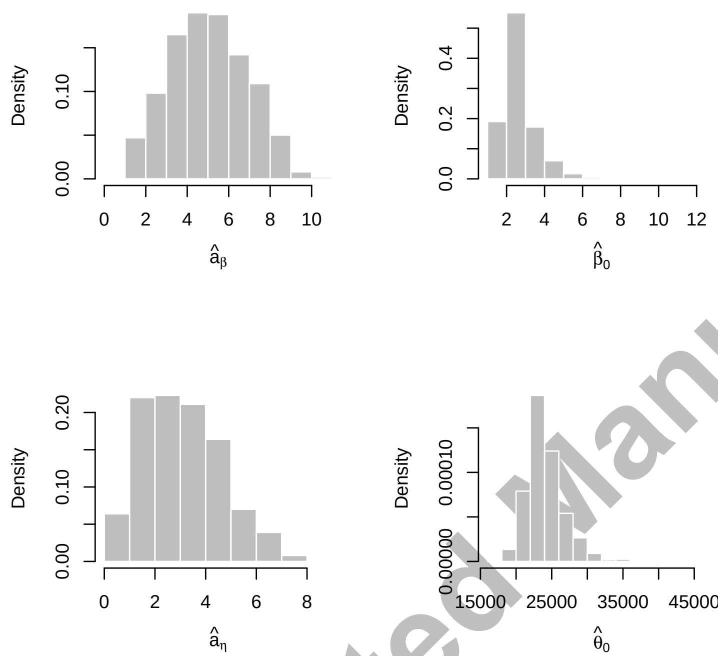

Figure 7 *Bootstrap sample histograms of $\hat{\boldsymbol{\phi}}$ based on $B = 1,000$ for the power transformers data.*

a quasi-conjugate posterior analysis. This reparameterization introduced a difficulty in the sense that, when the truncation times are different, it is unreasonable to assume exchangeability in the second stage prior distribution. A parametric empirical Bayes approach was carried out in order to estimate the model parameters. The hyperpameter vector $\boldsymbol{\phi}$ was estimated by maximizing its posterior density, or equivalently, a marginal penalized likelihood function. Once that the hyperparameters were estimated, approximations to the estimates of the system specific parameters were obtained using an *iid* Monte Carlo sample from $p(\boldsymbol{\beta}, \boldsymbol{\eta}|\boldsymbol{D}, \hat{\boldsymbol{\phi}})$. This Monte Carlo sample can be obtained using a simple and efficient rejection sampling algorithm. Furthermore, a parametric bootstrap method was used to correct the standard deviations of point estimates and the HPD intervals by taking into account the uncertainty in the estimate of the hyperparameters. These methods were used to analyze a real data set regarding failure times of 40 power transformers, including estimation of the optimal preventive maintenance time

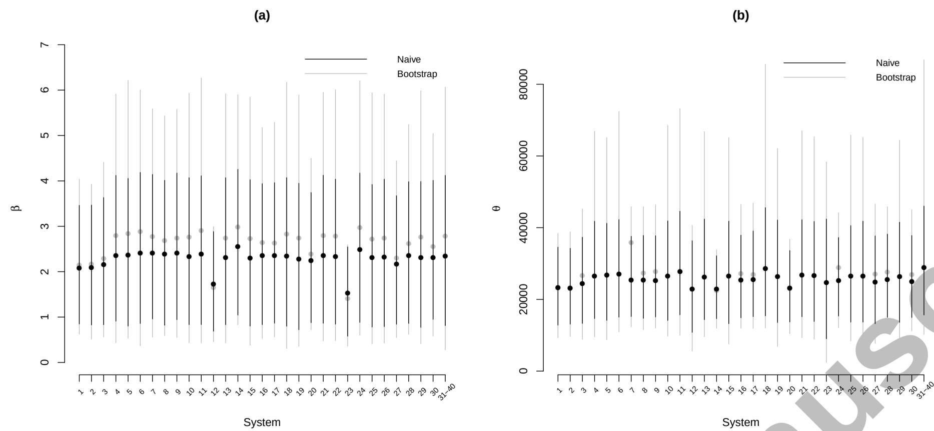

Figure 8 *Naive and bootstrap PEB 95% HPD credible intervals of the parameters β_i* ***(a)****, and θ_i* ***(b)****. The points are posterior expectations.*

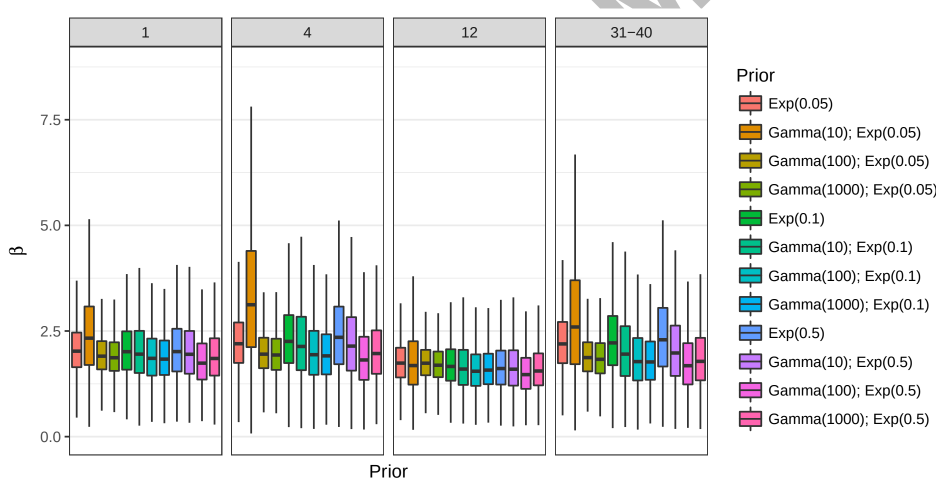

Figure 9 *Sensitivity analysis. Boxplot of β_i for systems 1, 4, 12 and 31-40. Prior* $\mathrm{Exp}(\cdot)$ *means exponential density priors for a_β and a_η, and uniform priors for β_0 and θ_0. Prior* $\mathrm{Gamma}(\cdot); \mathrm{Exp}(\cdot)$ *means exponential density priors for a_β and a_η, and gamma priors for β_0 and θ_0.*

considering block policy.

A fully Bayesian hierarchical model (BHM) could be viewed as an alternative approach for estimation of the parameters of the hierarchical PLP model. However the implementation of BHM generally requires the implementation of Markov Chain Monte Carlo methods. These methods involves

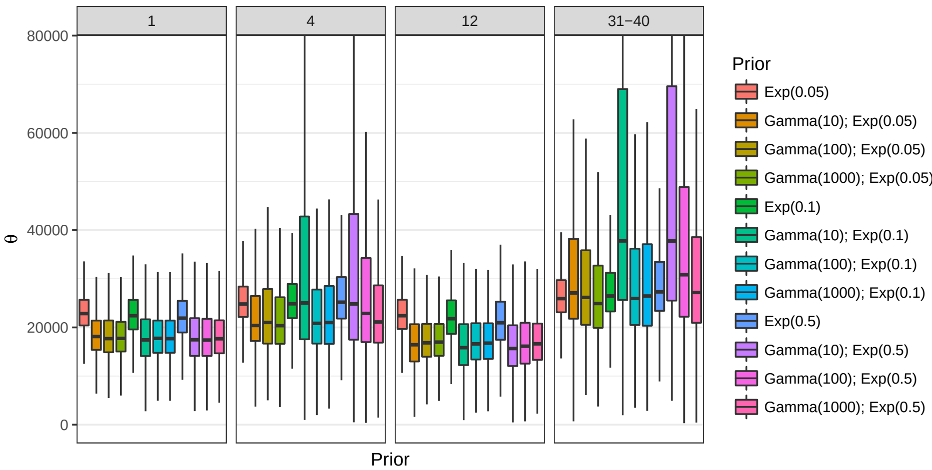

Figure 10 *Sensitivity analysis. Boxplot of θ_i for systems 1, 4, 12 and 31-40. Prior* $\mathrm{Exp}(\cdot)$ *means exponential density priors for a_β and a_η, and uniform priors for β_0 and θ_0. Prior* $\mathrm{Gamma}(\cdot); \mathrm{Exp}(\cdot)$ *means exponential density priors for a_β and a_η, and gamma priors for β_0 and θ_0.*

the specification of fine-tuning parameters and checking of chain convergence, which could not be trivial for researchers in the field of relibility of repairable systems (e.g. engineers, managers, economists, etc.). We believe that the suggested PEB approach avoid these potential complicators.

Acknowledgments

This work was partially supported by CNPq, CAPES, FAPEMIG and FAPDF.

Appendix A: Starting values for the maximum posterior estimation

The main ideia is to use the ML estimates of β_i and η_i as the true values in the second stage prior (2.2). Let $\hat{\boldsymbol{\beta}}_{ML}$ and $\hat{\boldsymbol{\eta}}_{ML}$ be the vectors of ML estimates for those systems with $n_i > 0$ (the ML estimate of β_i does not exist when $n_i > 0$). Taking logarithms in (2.2) and replacing the actual β_i

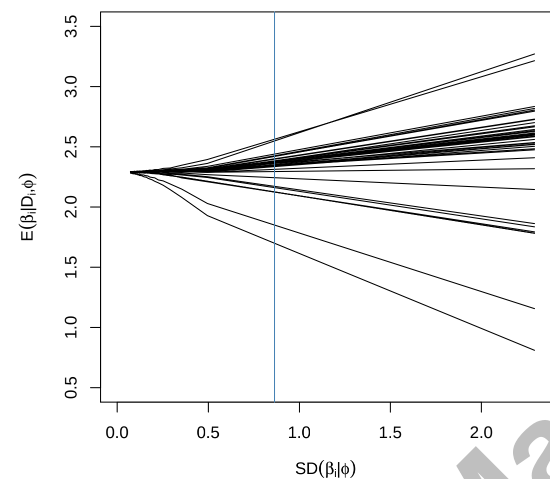

Figure 11 *Posterior means of β_i for the power transformers data, as a function of the prior standard deviation $SD(\beta_i|\phi)$, conditionally on $\hat{a}_\eta = 5.28$, $\hat{\theta}_0 = 18,399.20$, $\hat{\beta}_0 = 2.29$ and a sequence of $a_\beta \in (1; 1,000)$. For each configuration value $(\hat{a}_\beta^{(b)}, \hat{\beta}_0^{(b)}, \hat{a}_\eta, \hat{\theta}_0)$, a sample of size $1,000$ of β_i was generated and the sample mean was computed. The blue vertical line is the observed prior standard deviation $\mathrm{SD}(\beta_i|\hat{\boldsymbol{\phi}}) = 0.87$.*

and η_i by their ML estimates we obtain

$$\begin{aligned}\log\, p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}) = \sum_{i:n_i>0} \Big\{ & a_\eta[\log{(a_\eta)} + \hat{\beta}_i \log(\theta_0/\tau_i)] - \log\Gamma(a_\eta) \\ &+ \ (a_\eta - 1)\log(\hat{\eta}_i) - \hat{\eta}_i a_\eta (\theta_0/\tau_i)^{\hat{\beta}_i} \\ &+ \ a_\beta \log(a_\beta/\beta_0) - \log\Gamma(a_\beta) \\ &+ \ (a_\beta - 1)\log(\hat{\beta}_i) - \hat{\beta}_i(a_\beta/\beta_0) \Big\}. \end{aligned} \qquad \text{(A.1)}$$

Hence, we take as starting values for $\boldsymbol{\phi}$ the solution of $\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))/\partial\boldsymbol{\phi} =$

$\mathbf{0}$, that is

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial a_\beta} = \sum_{i:n_i>0} \left[\log(a_\beta/\beta_0) + 1 - \psi(a_\beta) + \log(\hat{\beta}_i) - \frac{\hat{\beta}_i}{\beta_0} \right] = 0, \tag{A.2}$$

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial \beta_0} = \frac{a_\beta}{\beta_0} \sum_{i:n_i>0} \left[\frac{\hat{\beta}_i}{\beta_0} - 1 \right] = 0, \tag{A.3}$$

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial a_\eta} = \sum_{i:n_i>0} \Big[\log(a_\eta) + 1 + \hat{\beta}_i \log(\theta_0/\tau_i) - \psi(a_\eta) + \log(\hat{\eta}_i) - \hat{\eta}_i (\theta_0/\tau_i)^{\hat{\beta}_i} \Big] = 0, \tag{A.4}$$

$$\frac{\partial \log(p(\hat{\boldsymbol{\beta}}_{ML}, \hat{\boldsymbol{\eta}}_{ML}|\boldsymbol{\phi}))}{\partial \theta_0} = \frac{a_\eta}{\theta_0} \sum_{i:n_i>0} \left[\hat{\beta}_i - \hat{\beta}_i \hat{\eta}_i \left(\frac{\theta_0}{\tau_i} \right)^{\hat{\beta}_i} \right] = 0. \tag{A.5}$$

Let K_* be the number of systems with $n_i > 0$. From Equation (A.3) we obtain that $\tilde{\beta}_0 = K_*^{-1} \sum_{i:n_i>0} \hat{\beta}_i$ and replacing β_0 by $\tilde{\beta}_0$ in Equation (A.2), we obtain $\tilde{a}_\beta$ as the solution of

$$\log(\tilde{a}_\beta) - \psi(\tilde{a}_\beta) - \log(\tilde{\beta}_0) - K_*^{-1} \sum_{i:n_i>0} \log(\hat{\beta}_i) = 0. \tag{A.6}$$

From Equation (A.5) we obtain that $\tilde{\theta}_0$ is the solution of

$$K_*^{-1} \sum_{i:n_i>0} \hat{\beta}_i - K_*^{-1} \sum_{i:n_i>0} \hat{\beta}_i \hat{\eta}_i (\tilde{\theta}_0/\tau_i)^{\hat{\beta}_i} = 0. \tag{A.7}$$

Finally, we replace θ_0 by $\tilde{\theta}_0$ in Equantion (A.4) to obtain $\tilde{a}_\eta$ as the solution of

$$\log(\tilde{a}_\eta) - \psi(\tilde{a}_\eta) - K_*^{-1} \sum_{i:n_i>0} \left[\hat{\beta}_i \log(\tilde{\theta}_0/\tau_i) + \log(\hat{\eta}_i) - \hat{\eta}_i (\tilde{\theta}_0/\tau_i)^{\hat{\beta}_i} \right] = 0. \tag{A.8}$$

We note that Equations (A.6) to (A.8) are all univariate and hence can be solved by simple numerical procedures. In the real data example analyzed in Section 4, these starting values were close to the final estimates.

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer.

Arab, A., Rigdon, S. E. and Basu, A. P. (2012). Bayesian Inference for the Piecewise Exponential Model for the Reliability of Multiple Repairable Systems. *Journal of Quality Technology* **44** 28-38.

Baker, R. D. (1996). Some New Tests of the Power Law Process. *Technometrics* **38** 256-265.

Barlow, R. and Hunter, L. (1960). Optimum Preventive Maintenance Policies. *Operations Research* **8** 90-100.

Bhattacharjee, M., Arjas, E. and Pulkkinen, U. (2003). Modelling Heterogeneity in Nuclear Power Plant Valve Failure Data. In *Mathematical and Statistical Methods for Reliability* (B. H. Lindqvist and K. A. Doksum, eds.) 341-353. World Scientific Publishing.

Carlin, B. P. and Gelfand, A. E. (1990). Approaches for the empirical Bayes confidence intervals. *Journal of the American Statistical Association* **85** 105-114.

Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician* **39** 83–87.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events. Statistics for Biology and Health*. Springer.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust Empirical Bayes Analyses of Event Rates. *Technometrics* **29** 1–15.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall, New York.

George, E. I., Makov, U. E. and Smith, A. F. M. (1993). Conjugate likelihood distributions. *Scandinavian Journal of Statistics* **20** 147–156.

Gilardoni, G. L. and Colosimo, E. A. (2007). Optimal Maintenance Time for Repairable Systems. *Journal of Quality Technology* **39** 48-53.

Gilardoni, G. L. and Colosimo, E. A. (2011). On the superposition of overlapping Poisson processes and nonparametric estimation of their intensity function. *Journal of Statistical Planning and Inference*.

Gilardoni, G. L., Oliveira, M. D. d. and Colosimo, E. A. (2013). Nonparametric estimation and Bootstrap confidence interval for the optimal maintenance time of a repairable system. *Computational Statistics & Data Analysis* **63** 113-124.

Giorgio, M., Guida, M. and Pulcini, G. (2014). Repairable system analysis in presence of covariates and random effects. *Reliability Engineering and System Safety* **131** 271-281.

Guida, M., Calabria, R. and Pulcini, G. (1989). Bayes inference for a non-homogeneous Poisson process with power intensity law. *IEEE Transactions on Reliability* **38** 603–609.

Guida, M. and Pulcini, G. (2005). Bayesian reliability assessment of repairable systems during multi-stage development programs. *IIE Transactions* **37** 1071 - 1081.

Hamada, M. S., Wilson, A. G., Reese, C. S. and Martz, H. F. (2008). *Bayesian Reliability. Springer Series in Statistics*. Springer.

Huang, Y. S. (2001). A decision model for deteriorating repairable systems. *IIE Transactions* **33** 479.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models). *Journal of the American Statistical Association* **84** 717-726.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes Confidence Intervals Based on Bootstrap Samples. *Journal of the American Statistical Association* **82** 739-750.

Lawless, J. F. (1987). Regression Methods for Poisson Process Data. *Journal of the American Statistical Association* **82** 808-815.

Lindqvist, B. H. (2006). On the Statistical Modeling and Analysis of Repairable Systems. *Statistical Science* **21** 532–551.

Lindqvist, B. H., Elvebakk, G. and Heggland, K. (2003). The Trend-Renewal Process for Statistical Analysis of Repairable Systems. *Technometrics* **45** 31-44.

Mazzuchi, T. A. and Soyer, R. (1996). A Bayesian perspective on some replacement strategies. *Reliability Engineering and System Safety* **51** 295-303.

Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association* **78** 47–55.

Oliveira, M. D., Colosimo, E. A. and Gilardoni, G. L. (2012). Bayesian inference for power law processes with applications in repairable systems. *Journal of Statistical Planning and Inference* **142** 1151–1160.

Pan, R. and Rigdon, S. E. (2009). Bayes Inference for General Repairable Systems. *Journal of Quality Technology* **41** 82-94.

Pérez, C. J., Martín, J. and Rufo, M. J. (2006). Sensitivity estimations for Bayesian inference models solved by MCMC methods. *Reliability Engineering and System Safety* **91** 1310-1314.

Perkins, S., Cohen, M., Rahme, E. and Bernatsky, S. (2012). Melanoma and rheumatoid arthritis (brief report). *Clinical Reumathology* **31** 1001-1003.

Pesaran, M. H., Pettenuzzo, D. and Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks. *The Review of Economic Studies* **73** 1057-1084.

Rigdon, S. E. and Basu, A. P. (2000). *Statistical Methods for the Reliability of Repairable Systems*. John Wiley & Sons.

Ryan, K. J., Hamada, M. S. and Reese, C. S. (2011). A Bayesian Hierarchical Power Law Process Model for Multiple Repairable Systems with an Application to Supercomputer Reliability. *Journal of Quality Technology* **43** 209-223.

Schilling, R. L. (2005). *Measures, Integrals and Martingales*. Cambridge University Press.

R Core Team (2013). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

R. C. P. Reis
E. A. Colosimo
Departamento de Estatística
Universidade Federal de Minas Gerais
Belo Horizonte, MG 31270-901
Brazil
E-mail: rcpr@est.mest.ufmg.br; enricoc@est.ufmg.br

G. L. Gilardoni
Departamento de Estatística
Universidade de Brasília, Brasília
DF 70910-900
Brazil
E-mail: gilardon@unb.br