Nonhomogeneous Poisson process with nonparametric frailty and covariates

Vaclav Slimacek[a,*], Bo Henry Lindqvist[a]

[a] *Norwegian University of Science and Technology, Trondheim, Norway*

Abstract

The nonhomogeneous Poisson process is commonly used in the modeling of failure times of complex repairable systems. In practice there may be a substantial heterogeneity in the failure behavior among apparently identical repairable systems. In this paper we introduce a new approach for statistical modeling of failures and the corresponding statistical inference when there is both an observable and unobservable heterogeneity between such systems. The approach is partly nonparametric and hence avoids making restrictive assumptions about the underlying process. The main feature of the approach is the elimination of the effect of unobservable heterogeneity, which leaves an optimization problem involving the observable covariates only. The new method is introduced in a power law process setting and can easily be extended to general nonhomogeneous Poisson process. The satisfactory performance of the method is verified in an extensive simulation study as well as in a case study, and the method compares favorably to the gamma frailty model and to the classical regression model not assuming an unobserved heterogeneity. The approach can be adapted for a wide class of models.

Keywords:
nonhomogeneous Poisson process, unobserved heterogeneity, nonparametric estimation, covariates, power law process, recurrent events

1. Introduction

The reliability of complex repairable systems is commonly studied by using counting process models, where events correspond to failures of the system. The most prominent model is the nonhomogeneous Poisson processes (NHPP), which is completely described by its rate of occurrence of failures (ROCOF). This description should take into account that in many real life applications there is a substantial heterogeneity between apparently identical repairable systems. The motivation for this paper was the joint presence of observed and unobserved heterogeneity in failure patterns of German onshore wind turbines analyzed in detail in [15]. Several factors like size of a turbine, manufacturer or local climatic conditions are known to influence the reliability of wind turbines ([5]). These may be represented by a set of covariates, leading to the consideration of regression models. Nevertheless, there are other

*Corresponding author
Email address: v.slimacek@gmail.com (Vaclav Slimacek)

factors that potentially may affect the reliability, e.g., position of a turbine inside a wind farm or specific technology used in a turbine, which are not recorded or are difficult to quantify. The latter types of factors are not necessarily observable, but may still have a significant effect on the reliability. Thus, if ignored, a regression analysis of repairable systems may give misleading or wrong results.

In [14] we introduced a new approach for modeling and analysis of repairable systems with an unobservable heterogeneity. We adopted the common approach of multiplication of the basic ROCOF by a positive random variable, a so called frailty, taking independent values across the systems. While in traditional frailty models it is necessary to make parametric assumptions about the distribution of the frailties, we made a nonparametric approach, thus avoiding numerical and other problems that may occur with the parametric models, especially in large data sets and when using complicated models.

In this paper we extend the technique from [14], where focus is now on regression models for NHPPs, involving observed covariates in addition to unobservable heterogeneity. The main idea is to eliminate the possible disturbances of the unobserved heterogeneity, thus enabling estimation of the net effect of the observable covariates in the behavior of the systems. The most common way to incorporate covariates into the ROCOF model is to multiply a baseline ROCOF by a function of the covariates, with the exponential function being the most common choice, leading to what corresponds to the proportional hazards model in survival analysis, which here also might be called the proportional ROCOF model. A heterogeneity can then be added to such a model by multiplication of the frailty, as indicated above. A classical reference for regression modeling in NHPPs is [10], where heterogeneity (called random effects) is also treated by assuming the standard approach of letting frailties be gamma distributed. A subsequent review of applications of proportional hazards models in repairable systems was given in [9]. More recent reviews are given in [12] and [7]. The former of these papers advocates the use of trend tests to guide in the search for appropriate failure models. The latter paper, on the other hand, considers the problem of model selection for multiple repairable units, with emphasis on unobservable heterogeneity, dependencies and trend.

It seems that the majority of analyses of repairable systems do not include factors or covariates that may affect the failure process. Thus the literature on repairable systems with covariates is much less rich than, for example, the literature on regression models for recurrent events in biostatistics. There are, however, several recent approaches in reliability where covariates play an important role. For example, [13] used covariates such as length or diameter of pipes, age and presence of clay in order to explain the reliability of pipe failures in water networks. The same data were reanalyzed in [1] using time-dependent covariates. Motivated by a similar application and using similar covariates, [11] introduced a dynamic NHPP, where the ROCOF of a system increases with the observed number of failures. The importance of taking covariates and differences in environment into consideration is furthermore discussed in the paper [2], analyzing data for drill bits in a bauxite mine. The paper [3] made use of the power law process for multiple repairable units with differing reliability characteristics to predict the expected number of failures for a fleet of military aircrafts, while [18] extended the power law model by introducing a single multiplicative covariate with the interpretation of a known scaling of the system's failure behavior. The paper [20] proposed the use of a proportional hazards model with time-varying covariates for residual life prediction of repairable systems. The paper [19] models operational and environmental stress factors as covariates in repairable systems. A different approach to involving both observable and unobservable heterogeneity has recently been presented in [8]. The authors' main idea is to introduce a joint distribution of the parameters of the baseline power law ROCOF function, and estimate its parameters by an empirical Bayes procedure.

The rest of this paper is organized as follows. In Section 2 we review the main results of [14]. Section 3 is the main section where we introduce the new model which extends the model in [14] by inclusion of covariates. In particular we present a novel approach for estimation of the coefficient vector which describes the influence of the covariates. Section 4 describes briefly a comprehensive simulation study and some of its conclusion, where the simulation study itself will be presented in the supplementary material. A case study using data from the WMEP database for German wind turbines [5] is presented in Section 5. Some concluding remarks are finally given in Section 6.

2. NHPP with covariates and nonparametric frailty

2.1. The case with unobserved heterogeneity and no covariates ([14])

In this section we review the main results from [14]. As in that article, consider a model where, conditionally on Z, the events (failures) follow an NHPP with ROCOF

$$\lambda(t|Z) = Zabt^{b-1} \tag{1}$$

where the power law basic intensity was chosen for simple illustration. Recall that Z is a positive random variable representing the frailty of the system under study. In order to avoid an identification problem concerning the scale parameter a we will here assume that $\mathrm{E}(Z) = 1$. Further, we assume that $\mathrm{Var}(Z)$ is finite.

Let m independent processes described by (1) be observed on the time interval $[0, \tau_j]$ for a random length $\tau_j > 0$ (realizations of a random variable τ), and let us denote number of observed events n_j with event times t_{ij} ($i = 1, \ldots, n_j$, $j = 1, \ldots, m$). Unobserved individual frailties (unobserved realizations of a random variable Z) will be denoted as $z_1, \ldots, z_m$. τ is assumed to be stochastically independent of Z.

In [14] we derived estimates of a, b and their variances, as well as predicted values for the individual frailties z_j. Also, estimates of $\mathrm{Var}(Z)$ were given. Thus, while standard models of this kind use parametric distributions for Z, e.g., the gamma distribution, the clue of [14] was to avoid any assumptions on the distribution of Z.

Below we review the estimators,

$$\hat{b} = \frac{\sum_{j=1}^{m} n_j - 1}{\sum_{j=1}^{m} n_j \log(\tau_j) - \sum_{j=1}^{m} \sum_{i=1}^{n_j} \log(t_{ij})} \tag{2}$$

$$\hat{a} = \frac{1}{m} \sum_{j=1}^{m} \frac{n_j}{\tau_j^{\hat{b}}} \tag{3}$$

$$\hat{z}_j = \frac{n_j}{\hat{a}\tau_j^{\hat{b}}} \tag{4}$$

Among several possible estimators, the preferred estimator of $\text{Var}(Z)$ in [14] was their equation (34),

$$\widehat{\text{Var}}(Z) = \widehat{\text{Var}}\left(\hat{z}_j\right) - \frac{\sum_{j=1}^{m} \tau_j^{-\hat{b}}}{m\hat{a}} \tag{5}$$

where $\widehat{\text{Var}}\left(\hat{z}_j\right)$ is the empirical variance of the z_j from (4).

2.2. *The new model: NHPP with covariates and nonparametric frailty*

Suppose now that for each system there is an observable covariate vector $\boldsymbol{X} = [X_1, \ldots, X_R]$. The idea is to modify the model from [14], as described in the previous section, so that the influence of $\boldsymbol{X}$ is taken into account. The model we shall consider for inclusion of $\boldsymbol{X}$ is of the Cox-type, where the conditional ROCOF, given the frailty Z and the covariate vector $\boldsymbol{X}$, can be written

$$\lambda(t|Z, \boldsymbol{X}) = Zabt^{b-1}\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right) \tag{6}$$

As before, $a > 0$ and $b > 0$ are unknown parameters, while $\boldsymbol{\beta}_0 = [\beta_{01}, \ldots, \beta_{0R}]$ is the vector of unknown coefficients ($'$ denotes the transposition of a vector). Z is as before a positive random variable with $E(Z) = 1$ and $\text{Var}(Z) < \infty$ representing the unobserved heterogeneity (frailty). Again, the process is observed from time 0 until a random time $\tau > 0$. The assumption is now that the variables Z, $\boldsymbol{X}$, τ are independent. Note, that the classical regression model (i.e., the model without unobserved heterogeneity) is included in this settings as a special case with $\text{P}(Z = 1) = 1$.

The basic properties of NHPPs imply that the conditional expectation and variance of the number of events N occurring within a time τ are given by, respectively,

$$\text{E}(N|Z, \boldsymbol{X}, \tau) = Za\tau^b\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right) \tag{7}$$

$$\text{Var}(N|Z, \boldsymbol{X}, \tau) = Za\tau^b\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right) \tag{8}$$

Let us assume that m independent processes of this type are observed, where the jth process is observed for a time τ_j and described by an observed vector of covariates $\boldsymbol{x}_j = [x_{j1}, \ldots, x_{jR}]$, and an unobserved individual frailty z_j. The number of observed events in the jth process is denoted as n_j and the observed event times in this process are denoted as t_{ij} ($i = 1, \ldots, n_j$, $j = 1, \ldots, m$).

The main interest is in the estimation of the coefficient vector $\boldsymbol{\beta}_0$ which explains the effect of covariates on the failure behavior of a process and which is discussed in the following section. The estimation of b, a and prediction of the z_j's and $\text{Var}(Z)$ are discussed in the section 2.4.

2.3. *Estimation of the effect of covariates*

The form of the conditional expected number of events N within a time τ given by (7) allows the construction of estimators of the coefficients $\boldsymbol{\beta}_0$ in (6). Taking the expected value of each side of (7) gives the unconditional expected number of events (with use of the assumed independence Z, τ and $\boldsymbol{X}$)

$$\text{E}(N) = \text{E}(\text{E}(N|Z, \boldsymbol{X}, \tau)) = \text{E}(Z)\,a\text{E}\left(\tau^b\right)\text{E}\left(\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right) \tag{9}$$

Similarly,

$$\mathrm{E}\left(\frac{N}{\tau^b \exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\right) = \mathrm{E}\left(\mathrm{E}\left(\frac{N}{\tau^b \exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\middle| Z, \boldsymbol{X}, \tau\right)\right) = \mathrm{E}(Z)\, a \tag{10}$$

Substitution of $\mathrm{E}(Z)\, a$ from equation (10) into equation (9) gives

$$\mathrm{E}\left(\frac{N}{\tau^b \exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\right) \frac{\mathrm{E}\left(\tau^b\right) \mathrm{E}\left(\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right)}{\mathrm{E}(N)} = 1 \tag{11}$$

Based on equation (11), let us define the function $h(\boldsymbol{\beta})$ by

$$h(\boldsymbol{\beta}) = \mathrm{E}\left(\frac{N}{\tau^b \exp\left(\boldsymbol{\beta}'\boldsymbol{X}\right)}\right) \frac{\mathrm{E}\left(\tau^b\right) \mathrm{E}\left(\exp\left(\boldsymbol{\beta}'\boldsymbol{X}\right)\right)}{\mathrm{E}(N)} - 1 \tag{12}$$

Equation (11) then states that the true value $\boldsymbol{\beta}_0$ lies in the set described as

$$\{\boldsymbol{\beta} : h(\boldsymbol{\beta}) = 0\} \tag{13}$$

With use of the double expectation rule, linearity of the expected value, equation (7) and assumed independencies we get

$$\begin{aligned}\mathrm{E}\left(\frac{N}{\tau^b \exp\left(\boldsymbol{\beta}'\boldsymbol{X}\right)}\right) &= \mathrm{E}\left(\mathrm{E}\left(\frac{N}{\tau^b \exp\left(\boldsymbol{\beta}'\boldsymbol{X}\right)}\middle| Z, \boldsymbol{X}, \tau\right)\right) = \mathrm{E}\left(\frac{1}{\tau^b \exp\left(\boldsymbol{\beta}'\boldsymbol{X}\right)} \mathrm{E}\left(N | Z, \boldsymbol{X}, \tau\right)\right) \\ &= \mathrm{E}(Z)\, a \mathrm{E}\left(\exp\left(\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right)'\boldsymbol{X}\right)\right)\end{aligned} \tag{14}$$

Thus the function $h(\boldsymbol{\beta})$ can be represented as (by use of equation (14) and equation (9))

$$h(\boldsymbol{\beta}) = \frac{\mathrm{E}\left(\exp\left(\boldsymbol{\beta}'\mathbf{X}\right)\right) \mathrm{E}\left(\exp\left(\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}\right)'\mathbf{X}\right)\right)}{\mathrm{E}\left(\exp\left(\boldsymbol{\beta}_0'\mathbf{X}\right)\right)} - 1 \tag{15}$$

As can be easily checked from equation (15), $h\left(\frac{\boldsymbol{\beta}_0}{2} + \boldsymbol{\beta}\right) = h\left(\frac{\boldsymbol{\beta}_0}{2} - \boldsymbol{\beta}\right)$ for all $\boldsymbol{\beta}$, i.e. $h(\boldsymbol{\beta})$ is symmetrical around the point $\boldsymbol{\beta} = \frac{\boldsymbol{\beta}_0}{2}$.

From the Cauchy-Schwarz inequality $|\mathrm{E}(UV)|^2 \leq \mathrm{E}\left(U^2\right) \mathrm{E}\left(V^2\right)$ with $U = \exp\left(\left(\frac{\boldsymbol{\beta}_0}{2} - \boldsymbol{\beta}\right)' \frac{\mathbf{X}}{2}\right)$, $V = \exp\left(\left(\frac{\boldsymbol{\beta}_0}{2} + \boldsymbol{\beta}\right)' \frac{\mathbf{X}}{2}\right)$ and $UV = \exp\left(\frac{\boldsymbol{\beta}_0}{2}' \mathbf{X}\right)$ it also follows that $h\left(\frac{\boldsymbol{\beta}_0}{2} \pm \boldsymbol{\beta}\right) > h\left(\frac{\boldsymbol{\beta}_0}{2}\right)$ for all $\boldsymbol{\beta} \neq \mathbf{0}$, i.e. the point $\boldsymbol{\beta} = \frac{\boldsymbol{\beta}_0}{2}$ is the unique minimum of the function $h(\boldsymbol{\beta})$. Thus we have the somewhat unexpected result

$$\boldsymbol{\beta}_0 = 2 \operatorname*{argmin}_{\boldsymbol{\beta}} \left(h(\boldsymbol{\beta})\right)$$

which is the key result in the estimation procedure for $\boldsymbol{\beta}_0$ to be presented below.

The idea is to minimize instead the sample version $\hat{h}(\boldsymbol{\beta})$ of $h(\boldsymbol{\beta})$ in (12), where the theoretical means are substituted by simple sample averages and parameter b is substituted by its estimator. The resulting estimator of $\boldsymbol{\beta}_0$ is then

$$\hat{\boldsymbol{\beta}} = 2 \operatorname*{argmin}_{\boldsymbol{\beta}} \left(\hat{h}(\boldsymbol{\beta})\right) = 2 \operatorname*{argmin}_{\boldsymbol{\beta}} \left(\frac{1}{m^2} \frac{\sum_{j=1}^{m} \tau_j^{\hat{b}}}{\sum_{j=1}^{m} n_j} \sum_{j=1}^{m} \frac{n_j}{\tau_j^{\hat{b}} \exp\left(\boldsymbol{\beta}'\boldsymbol{x}_j\right)} \sum_{k=1}^{m} \exp\left(\boldsymbol{\beta}'\boldsymbol{x}_k\right)\right) \tag{16}$$

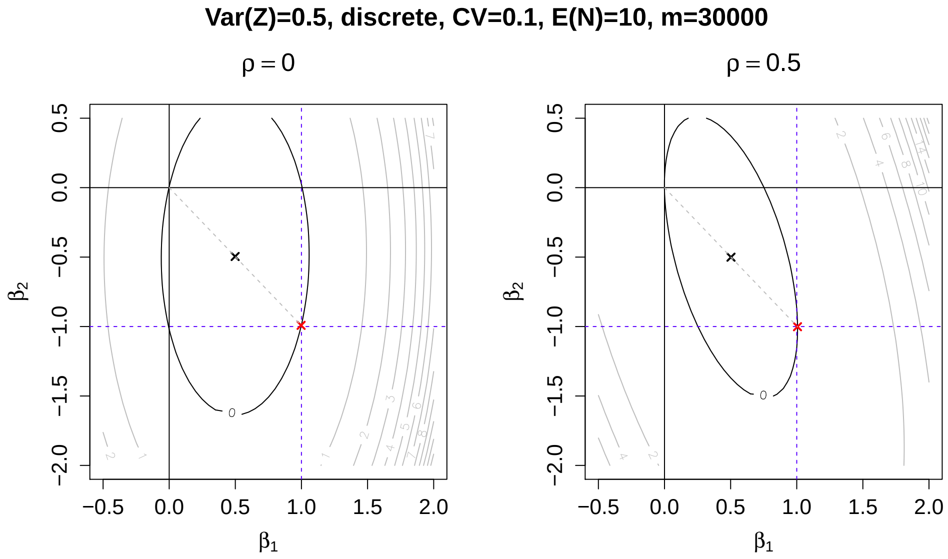

Figure 1: The gray contours represent the contours of the function $h(\boldsymbol{\beta})$ for the case $R = 2$ and selected values of the parameters and specific distributions of $\boldsymbol{X}$. The black contour represents the contour with $h(\boldsymbol{\beta}) = 0$. The black cross represents the minimum of $h(\boldsymbol{\beta})$, the red cross represents the estimate $\hat{\boldsymbol{\beta}}$, the intersection of the blue dashed lines represent the true coefficient vector $\boldsymbol{\beta}_0 = (1, -1)$. The curves are obtained by utilizing the consistency of $\hat{\boldsymbol{\beta}}$ and using the high value of $m = 30000$ to obtain curves that are for practical purposes equal to the theoretical ones.

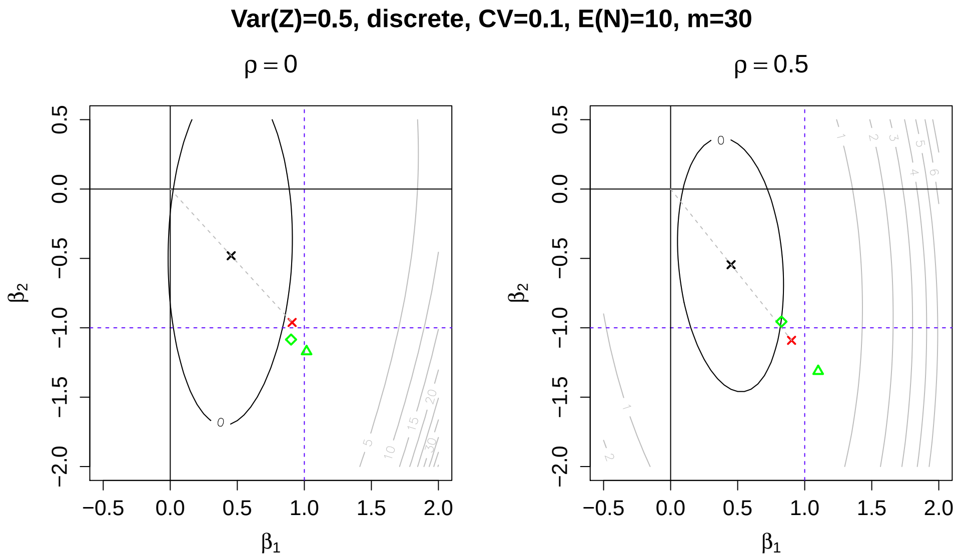

Figure 2: The underlying distributions are the same as for Figure 1, but the curves are based on a simulated sample of observations from $m = 30$ processes. The gray contours represent the contours of the empirical function $\hat{h}(\boldsymbol{\beta})$ with black contour representing the contour with value equal to 0. The black cross represents the minimum of $\hat{h}(\boldsymbol{\beta})$, the red cross represents the estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ which is represented by the intersection of the blue dashed lines. The green triangle represents the value estimated by the classical regression model which does not assume heterogeneity, while the green diamond represents the value estimated by gamma frailty model..

An illustration of the process of obtaining $\hat{\boldsymbol{\beta}}$ in (16) is presented in Figures 1 and 2, which are based on values of parameters used in the simulation study (see Section 3 for an explanation).

It can be shown that the estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$, i.e., $\hat{\boldsymbol{\beta}}$ converges (in probability) to $\boldsymbol{\beta}_0$ as m tends to infinity. The clue is that $h(\boldsymbol{\beta})$ is a convex function of $\boldsymbol{\beta}$, and the result then follows from Theorem 1, page 49, in [4].

Since the derived estimator $\hat{\boldsymbol{\beta}}$ is twice the minimum of the sample version of the function $h(\boldsymbol{\beta})$, the variance of the estimator $\hat{\boldsymbol{\beta}}$ can be estimated by adaptation of the approach used in [6], i.e. by using a Taylor expansion and the implicit function theorem, as briefly summarized in Appendix A.

2.4. *Estimation of other parameters*

2.4.1. *Estimation of b*

As was shown above, an estimate of b is needed in the estimation of $\boldsymbol{\beta}_0$. It is seen from the derivation in Appendix A of [14] that the problem of estimating b is not changed by the inclusion of covariates. In fact, in model (6) we can define the unobserved variable $\tilde{Z} = Z\exp(\boldsymbol{\beta}_0'\boldsymbol{X})$, which brings (6) on the same form as (1). Hence (2) (and the corresponding estimator of variance) is still valid.

2.4.2. *Estimation of a*

The estimation of a can be done in a way similar to the derivation of $\hat{a}$ for the case without covariates (see (3)). Since $E(Z) = 1$, the estimator of a can be defined by the sample version of equation (10), involving the already derived estimators, i.e.,

$$\hat{a} = \frac{1}{m}\sum_{j=1}^{m}\frac{n_j}{\tau_j^{\hat{b}}\exp\left(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_j\right)} \tag{17}$$

Alternatively, equation (9) could be used to construct an estimator of a.

The variance of $\hat{a}$ (given $\boldsymbol{x}_j$'s and τ_j's , $j = 1, \ldots, m$) can be derived with use of Taylor expansion and properties of the estimators $\hat{b}$ and $\hat{\boldsymbol{\beta}}$ and NHPPs and is briefly summarized in Appendix B.

2.4.3. *Estimation of unobserved heterogeneity*

The traditional frailty models are characterized by the variance of the unobserved effects which can be estimated with use of the properties of NHPPs following the recipes given by [14].

The unconditional variance of N can be computed as follows, with use of (7) and (8) and the assumption $\mathrm{E}(Z) = 1$.

$$\begin{aligned}\mathrm{Var}(N) &= \mathrm{E}(\mathrm{Var}(N|Z,\boldsymbol{X},\tau)) + \mathrm{Var}(\mathrm{E}(N|Z,\boldsymbol{X},\tau)) \\ &= a\mathrm{E}\left(\tau^b\right)\mathrm{E}\left(\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right) + a^2\left(\mathrm{E}\left(Z^2\right)\mathrm{E}\left(\tau^{2b}\right)\mathrm{E}\left(\exp\left(2\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right) - \mathrm{E}\left(\tau^b\right)^2\mathrm{E}\left(\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right)^2\right)\end{aligned} \tag{18}$$

which together with (9) gives

$$\mathrm{Var}(Z) = \frac{\mathrm{E}\left(N^2\right) - \mathrm{E}(N)}{\mathrm{E}(N)^2}\frac{\mathrm{E}\left(\tau^b\right)^2}{\mathrm{E}\left(\tau^{2b}\right)}\frac{\mathrm{E}\left(\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right)^2}{\mathrm{E}\left(\exp\left(2\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right)} - 1 \tag{19}$$

Using empirical forms of the expectations and substituting the estimators of the parameter b and the vector $\boldsymbol{\beta}$ in (19) gives the estimator $\widehat{\mathrm{Var}}(Z)$, i.e.,

$$\widehat{\mathrm{Var}}(Z) = \frac{1}{m}\frac{\sum_{j=1}^{m} n_j^2 - \sum_{j=1}^{m} n_j}{\left(\sum_{j=1}^{m} n_j\right)^2}\frac{\left(\sum_{j=1}^{m}\tau_j^{\hat{b}}\right)^2}{\sum_{j=1}^{m}\tau_j^{2\hat{b}}}\frac{\left(\sum_{j=1}^{m}\exp\left(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_j\right)\right)^2}{\sum_{j=1}^{m}\exp\left(2\hat{\boldsymbol{\beta}}'\boldsymbol{x}_j\right)} - 1 \tag{20}$$

The method of moments using equation (7) together with equation (10) allows us to express the individual frailties z_j as

$$z_j = \frac{n_j}{a\tau_j^b \exp\left(\boldsymbol{\beta}_0'\boldsymbol{x}_j\right)} \tag{21}$$

This expression can also be obtained by the likelihood approach considered in [14] in which the individual frailties z_j are viewed as parameters. Using the derived estimators in (21) defines the estimators of the individual frailties $\hat{z}_j$, $(j = 1, \ldots, m)$

$$\hat{z}_j = \frac{n_j}{\hat{a}\tau_j^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_j\right)} \tag{22}$$

Note, that the empirical mean of the estimated individual frailties is equal 1, i.e. $\frac{1}{m}\sum_{j=1}^{m}\hat{z}_j = 1$.

A natural choice for the estimator of the variance of the unobserved effects is the empirical variance of the estimated $\hat{z}_j$'s, which will be denoted as $\widehat{\mathrm{Var}}\left(\hat{z}_j\right)$. These are estimates of the variance $\mathrm{Var}\left(\frac{N}{a\tau^b\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\right)$, which can be computed as

$$\begin{aligned}\mathrm{Var}\left(\frac{N}{a\tau^b\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\right) &= \mathrm{E}\left(\mathrm{Var}\left(\frac{N}{a\tau^b\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\middle| Z,\tau,\boldsymbol{X}\right)\right) + \mathrm{Var}\left(\mathrm{E}\left(\frac{N}{a\tau^b\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right)}\middle| Z,\tau,\boldsymbol{X}\right)\right)\\ &= \frac{\mathrm{E}\left(\tau^{-b}\right)\mathrm{E}\left(\exp\left(-\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right)}{a} + \mathrm{Var}(Z)\end{aligned} \tag{23}$$

This means that $\widehat{\mathrm{Var}}\left(\hat{z}_j\right)$ overestimates the true variance of the unobserved effects by the factor $\frac{\mathrm{E}\left(\tau^{-b}\right)\mathrm{E}\left(\exp\left(-\boldsymbol{\beta}_0'\boldsymbol{X}\right)\right)}{a}$. The formula (23) hence suggests an estimator of $\mathrm{Var}(Z)$,

$$\widehat{\mathrm{Var}}(Z) = \widehat{\mathrm{Var}}\left(\hat{z}_j\right) - \frac{\sum_{j=1}^{m}\tau_j^{-\hat{b}}\sum_{j=1}^{m}\exp\left(-\hat{\boldsymbol{\beta}}'\boldsymbol{x}_j\right)}{m^2\hat{a}} \tag{24}$$

which generalizes (5) to the case which includes observed covariates.

2.5. Remarks

- Since $h(\mathbf{0}) = 0$ by (15), the sample version $\hat{h}(\mathbf{0})$ goes through 0 asymptotically.
- The estimation of $\boldsymbol{\beta}$ using minimization of $\hat{h}(\boldsymbol{\beta})$ derived in this paper can be used also for models without unobserved heterogeneity.

- From the sample version $\hat{h}(\boldsymbol{\beta})$ is seen that the minimized function is a strictly convex, smooth and continuous function with one global unique minimum (no other local minima or maxima). Therefore, the numerical minimization of this function is stable, robust and quick.
- Efficient algorithms for the computation of the covariance matrix of $\hat{\boldsymbol{\beta}}$ exist ([6]).
- Since the estimation method mainly uses analytical expressions and simple numerical minimization, it is generally very quick. Bootstrapping can also conveniently be used for the estimation of the variance of the estimated parameters.
- The estimation process depends heavily on the estimation of means. It is, therefore, sensitive to extreme observations and outliers and a robust method for the estimation of the mean might be beneficial.
- A function $\tilde{h}(\boldsymbol{\beta})$ analogous to the function $h(\boldsymbol{\beta})$ which does not involve time can also be derived,

  $$\tilde{h}(\boldsymbol{\beta}) = \mathrm{E}\left(\frac{N}{\exp(\boldsymbol{\beta}'\boldsymbol{X})}\right)\frac{\mathrm{E}(\exp(\boldsymbol{\beta}'\boldsymbol{X}))}{\mathrm{E}(N)} - 1 \tag{25}$$

  Estimators of $\boldsymbol{\beta}_0$ based on the function $\tilde{h}(\boldsymbol{\beta})$ will have larger variance than estimators based on $h(\boldsymbol{\beta})$.
- Since the estimation of a, b and $\mathrm{Var}(Z)$ generalize the approach of [14], the remarks herein are valid also in that case.
- The derived estimation process can easily be generalized to other parametric NHPP models, i.e., for processes with ROCOF

  $$\lambda(t|Z,\boldsymbol{X}) = Z\lambda_0(t,\boldsymbol{\theta})\exp\left(\boldsymbol{\beta}_0'\boldsymbol{X}\right) \tag{26}$$

  where $\lambda_0(\boldsymbol{\theta}, t)$ denotes a basic ROCOF described by the parameters $\boldsymbol{\theta}$. The estimation process will then be an extension of the approach sketched in Section 3 of [14].

3. Simulation study

The functionality of the derived approach was tested in a simulation study which follows the setting of the simulation study in the previous paper [14]. Comparisons were made to the approaches using the gamma frailty model and the classical model without unobserved heterogeneity.

The aim of the simulation study was to investigate the dependence on the number of systems m; expected number of events per system $\mathrm{E}(N)$; the coefficient of variation (CV = ratio of standard deviation and mean) of the distribution of observation time τ (modeled by lognormal distribution), correlation between covariates; and presence of unobserved heterogeneity. The power law process with ROCOF $\lambda_0(t|a,b) = abt^{b-1}$ was chosen as the basic ROCOF, with $a = 1$ and $b = 1.3$.

Four different values of m were considered, $m = 10, 30, 100, 300$. Parameters were, furthermore, adjusted to given values of $\mathrm{E}(N) = 3, 10, 30, 100$ and $CV = 0.1$ or 0.5. Two cases with unobserved heterogeneity and one case without unobserved heterogeneity were considered - *discrete frailty*, where frailty is modeled as two unobserved groups in data (each system belongs to one of the groups with probability $\frac{1}{2}$, and $\mathrm{Var}\left(\frac{Z}{\mathrm{E}(Z)}\right) = 0.5$); *gamma frailty*, where frailty is gamma distributed with mean equal to 1 and variance equal to 0.5; and *no frailty*. The covariates x_1 and x_2 were chosen as bivariate normal with mean 0, standard deviations $\sigma_1 = 1$ and $\sigma_2 = 0.5$, and correlation ρ equal to 0 or 0.5. The values for the corresponding coefficients were chosen as $\beta_1 = 1$ and $\beta_2 = -1$.

The detailed setup of the simulation study is described in the supplementary material.

For each combination of ρ, m, $\mathrm{E}(N)$, CV, and each case of unobserved heterogeneity, 10000 simulations were run and processed. The model introduced in this paper was fitted to each of the simulated datasets, as well as the gamma frailty model and the classical model without unobserved heterogeneity. The unknown parameters a, b, $\boldsymbol{\beta}_0$, $\mathrm{Var}(Z)$ (when appropriate) were estimated for each model and simulated dataset.

The classical model and the gamma frailty model were fitted by numerical maximization of the loglikelihoods (displayed in the supplementary material).

3.1. Summary of the results of the simulation study

- *Empirical formulas*

  The empirical results from the simulation study agree satisfactorily with the theoretically derived formulas.

- *Effect of covariates*

  The detailed results of the estimation of parameters β_1 and β_2 can be found in the supplementary material in the tables 3, 4, 5, 6, 7 and 8. The results confirm the validity of the estimation approach introduced in this paper.

  Comparing the mean squared errors, the classical model has the highest values in the simulations with unobserved heterogeneity, while in the simulations without unobserved heterogeneity it is the best model (as expected since it is the true model in this case). The estimation based on the new approach is comparable to the results obtained from the gamma frailty model in the simulations with unobserved heterogeneity, although the gamma frailty model has slightly lower mean squared errors in the simulations with unobserved heterogeneity modeled as gamma frailty. In the simulations without unobserved heterogeneity the gamma frailty model is almost identical to the classical model, while the estimation based on the new approach has slightly higher mean squared errors.

  The formula for the computation of the standard errors of the estimators of β_1 and β_2 given by (A.10) contains an estimator of the variance of the unobserved effects. Therefore the estimation of the standard errors of these parameters is strongly influenced by the quality of the estimation of the variance of the unobserved effects. If the variance of the unobserved effects is estimated well, then the theoretical standard errors are in a good agreement with the empirical results while in case of bad estimation of the variance of the unobserved effects the formula (A.10) can return negative values.

The standard errors of the estimators of the parameters β_1 and β_2, computed from the hessians of the loglikelihood functions, behave similarly as for the parameters a and b in the analogous simulation study in [14]. That means that they underestimate the empirical standard errors of these estimators in the simulations with unobserved heterogeneity in the classical model, while in the simulations without unobserved heterogeneity the standard errors computed from the hessians of the loglikelihood functions slightly overestimate the empirical standard errors in the classical model. The standard errors of the estimators of the parameters β_1 and β_2 in the gamma frailty model computed from the hessians of the loglikelihood functions slightly overestimate the empirical standard errors of the estimators β_1 and β_2.

- *Parameters a, b and $Var(Z)$*

  The results from the theoretical formulas describing the approach of this paper are in very good agreement with the empirical results, which confirms the validity of the derived formulas.

  The results of the estimation of the parameter a and b and $\mathrm{Var}(Z)$ agree with the results from the analogous simulation study in [14], and can be found in the supplementary material.

Remark related to the numerical solution

The optimizations were performed using the optimization procedure `optim()` in `R`.

It was found that the numerical maximizations of the loglikelihoods for the classical model and gamma frailty model are quite sensitive to the choice of starting points used by the optimization procedure (especially in case of fitting gamma frailty model to data without unobserved heterogeneity) while the finding of the minimum of the function defined by (16) was numerically stable (since, as was argued above, it is a strictly convex function with one global minimum). Moreover, it was observed that the estimation with use of the function (16) and related equations was generally quicker than finding the maximums of the loglikelihoods. The gain in speed depends on the choice of the starting points.

The problems with optimization of loglikelihood functions can of course be influenced by the implementations of the optimization procedures, which may be improved. On the other hand, it is necessary to be aware of these problems when working with real data.

4. Case study

Failure data from WMEP database which contains reports about failure stops of German onshore wind turbines between years 1989-2009 were analyzed as a real life example. More about this database can be found in [5] and detailed analysis of this dataset can be found in [15].

In total, 9900 failure stops from 702 wind turbines from 369 different wind farms were processed (most of wind farms consists of 1 turbine, the biggest wind farm consist of 28 turbines). The mean time of observation of a turbine was 10.8 years, on average 1.3 failures per turbine per year were observed with mean time between failures 260 days.

In order to illustrate the method of the present paper we considered two covariates, x_1 representing the rated power of a turbine (in MW), and covariate x_2 representing the local harshness of environment. These are well known from the literature to influence the reliability of wind turbines, see e.g. [17] and [16], and they were also found to be the most important covariates in the analyses of [15]. Note that, due to lack of information in the database about local climatic conditions, a proxy covariate for harshness of environment was constructed in [15] using the available information on number of stops caused by external natural factors such as lightning, high wind and icing. More precisely, the corresponding covariate, which is x_2 in the present study, was computed for each wind farm as the average number of such stops per year for the turbines of the farm. It is important to note that the corresponding turbine stops were then not considered as failures in the analysis.

The distribution of the two covariates can be seen in Figure 3. The transformed variables, $\log(x_1)$ and $\sqrt{x_2}$, were used in the analysis in order to avoid too big influence of a few large values.

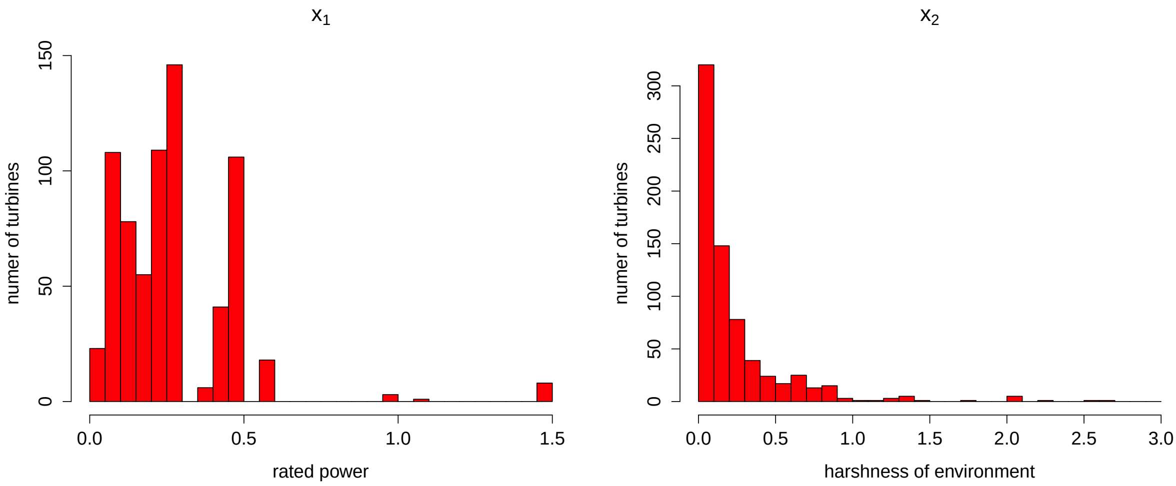

Figure 3: Distribution of the covariates x_1 (left) and x_2 (right) in the case study.

We fitted the nonparametric frailty model with covariates introduced in this paper, based on the power law process, to the data. The results are shown in table 1, together with the corresponding results when using the gamma frailty model. For comparison, we also fitted the classical power law model which does not take into account unobserved heterogeneity.

It is seen from table 1 that the results from the nonparametric model are in good agreement with the results from the gamma frailty model, which indicates a satisfactory behavior of our method. Since there is obviously a non-significant heterogeneity present in the data (revealed through the estimated $\mathrm{Var}(Z)$ and the corresponding low

Table 1: Result of fitting considered models to wind turbine failure data. $\widehat{\text{Var}}(Z)$ in the nonparametric frailty model is based on (24) and its standard error was estimated with use of bootstrapping (over wind turbines).

	$\hat{a}$	$SE(\hat{a})$	$\hat{b}$	$SE(\hat{b})$	$\hat{\beta}_1$	$SE(\hat{\beta}_1)$	$\hat{\beta}_2$	$SE(\hat{\beta}_2)$	$\widehat{\text{Var}}(Z)$	$SE(\widehat{\text{Var}}(Z))$
nonparametric frailty model	1.8042	0.1307	0.9418	0.0095	0.4303	0.0345	0.9060	0.0857	0.3162	0.0289
gamma frailty model	1.7375	0.1167	0.9429	0.0095	0.4164	0.0316	0.9442	0.0839	0.2998	0.0203
classical model	1.8549	0.0661	0.9471	0.0094	0.4378	0.0152	0.8521	0.0312	-	-

standard errors), the classical model, which does not involve heterogeneity, should be rejected as a good description of the data. As an observation from table 1, the estimated standard errors (with use of the hessian of the loglikelihood) of the parameters are smaller for the classical model, which apparently is caused by the more restricted assumptions of this model. The standard errors in the classical model were also computed with use of bootstrapping (over wind turbines which preserves the unobserved heterogeneity) and the results, which are summarized in table 2, show that the standard errors computed with use of hessian of the loglikelihood underestimate the real standard errors if the unobserved heterogeneity is not taken into account (the standard errors computed empirically and with use of hessian of the loglikelihoods are in good agreement for the gamma frailty model and the new method, as can be seen from the results of the simulation study). A consequence is that by constructing confidence intervals for the parameters by the standard method 'estimate plus two standard errors' we would obtain too optimistic intervals when using the classical (wrong) model.

Table 2: Comparison of standard errors computed with use of the hessian of the loglikelihood describing the classical model without unobserved heterogeneity and with use of bootstrapping (over wind turbines) using the same model.

	$SE(\hat{a})$	$SE(\hat{b})$	$SE(\hat{\beta}_1)$	$SE(\hat{\beta}_2)$
standard errors computed with use of hessian	0.0661	0.0094	0.0152	0.0312
bootstrap standard errors	0.1385	0.0125	0.0393	0.0709

One may then ask, what is the engineering conclusion that can be drawn from the case study? The estimated b is close to 1, which indicates that the failure process of individual wind turbines are approximately homogeneous Poisson processes, though with differing ROCOF's, due to differences in observed covariates as well as possibly with differing frailties Z.

Referring back to Section 2.3, Figure 4 illustrates the process of estimation of the regression parameters β_1 and β_2 corresponding to the observed covariates, and also shows the estimated parameters by each of the three methods considered. Since the estimated values of the coefficients β_1 and β_2 are positive and they have exponential influence on the ROCOF, the results confirm the negative effect of rated power and weather conditions on the number of failures.

Figure 5 plots the estimated individual frailties for both the gamma frailty model and the new proposed model. The latter are given by (22), while the gamma frailty model is considered in more detail in the supplementary material and in [14]. The figure shows that the frailties in the gamma frailty model are pushed more towards 1 compared to the ones obtained from the nonparametric model. For an explanation of this fact, see the discussion of the gamma frailty model in Section 2.2 of [14].

The estimated frailties can be viewed as coming from latent covariates which were not observed, but which if observed would have had an effect on an individual system's ROCOF. Such unobserved ”covariates” might in the present case be effects such as differing maintenance strategies, differences in the manufacturing process, position of turbines inside a wind farm etc.

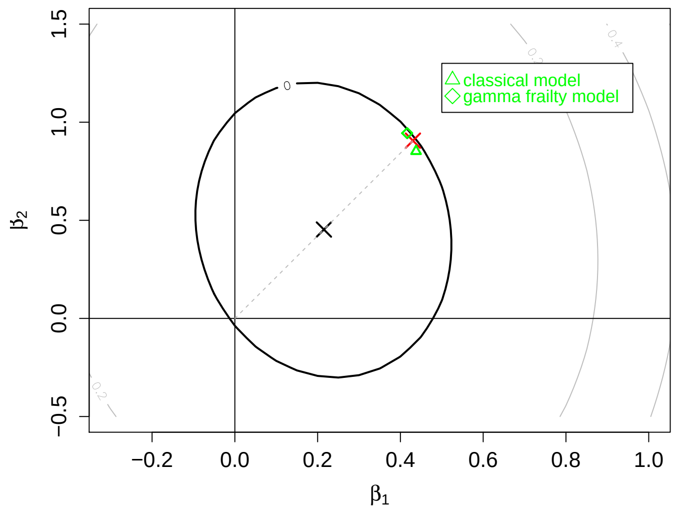

Figure 4: Illustration of the process of estimation of the parameters $\boldsymbol{\beta}_0$ introduced in this paper, i.e. by the equation (16), in the case study. The gray contours represent the contours of the function $h(\boldsymbol{\beta})$ with black contour representing the contour with value equal to 0. The black cross represents the minimum of the empirical version of the function $h(\boldsymbol{\beta})$, the red cross represents the estimated value. The green triangle represents the value estimated by classical model while the green diamond represents the value estimated by gamma frailty model.

5. Conclusions

This paper introduces a new method for estimation of the effect of covariates for a large class of models. The method extends the method introduced in [14] to the case with observable heterogeneity described by covariates. The main advantage of the new approach is that it is partly nonparametric and hence avoids making restrictive assumptions about the underlying process.

The functionality, advantages and disadvantages of the method, with respect to varying process parameters, have been illustrated and discussed in detail and compared to the classical models in a simulation study using the power law process as the basic model. In addition, the method has been applied in a case study using real data for wind turbine failures.

The simulation study shows that the new method performs very competitively compared to the classical model and to the gamma frailty model, while avoiding restrictive assumptions and numerical problems related to the use of the gamma frailty model. The method appears to be very well suited for solution by numerical methods.

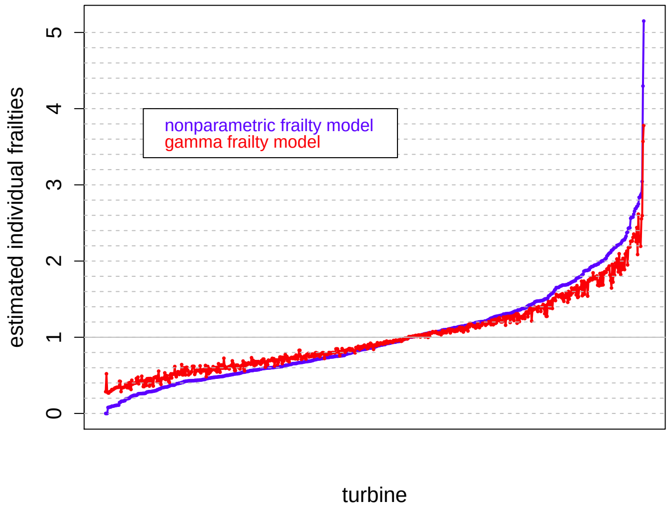

Figure 5: Estimated individual frailties (for each wind turbine) in the case study by the nonparametric frailty model (blue) and by the gamma frailty model (red). Turbines are ordered by the size of the estimated unobserved individual frailties in the nonparametric frailty model.

References

[1] Arasan, J., Ehsani, S., 2011. Modeling repairable system failures with repair effect and time dependent covariates. IAENG International Journal of Applied Mathematics 41 (3), 218–222.

[2] Barabadi, A., Barabady, J., Markeset, T., 2014. Application of reliability models with covariates in spare part prediction and optimization–a case study. Reliability Engineering & System Safety 123, 1–7.

[3] Block, J., Ahmadi, A., Tyberg, T., Kumar, U., 2014. Fleet-level reliability of multiple repairable units: A parametric approach using the power law process. International Journal of Performability Engineering 10 (3), 239–250.

[4] Dodge, Y., 2002. Statistical Data Analysis Based on the L B1 S-norm and Related Methods. Springer Science & Business Media.

[5] Faulstich, S., Hahn, B., Tavner, P. J., 2011. Wind turbine downtime and its importance for offshore deployment. Wind Energy 14 (3), 327–337.

[6] Fessler, J. A., 1996. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. Image Processing, IEEE Transactions on 5 (3), 493–506.

[7] Garmabaki, A., Ahmadi, A., Block, J., Pham, H., Kumar, U., 2016. A reliability decision framework for multiple repairable units. Reliability Engineering & System Safety 150, 78–88.

[8] Giorgio, M., Guida, M., Pulcini, G., 2014. Repairable system analysis in presence of covariates and random effects. Reliability Engineering & System Safety 131, 271–281.

[9] Kumar, D., 1995. Proportional hazards modelling of repairable systems. Quality and Reliability Engineering International 11 (5), 361–369.

[10] Lawless, J. F., 1987. Regression methods for poisson process data. Journal of the American Statistical Association 82 (399), 808–815.

[11] Le Gat, Y., 2014. Extending the yule process to model recurrent pipe failures in water supply networks. Urban Water Journal 11 (8), 617–630.

[12] Louit, D. M., Pascual, R., Jardine, A. K., 2009. A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data. Reliability Engineering & System Safety 94 (10), 1618–1628.

[13] Røstum, J., 2000. Statistical modelling of pipe failures in water networks. Tech. rep., PhD Dissertation, Norwegian University of Science and Technology.

[14] Slimacek, V., Lindqvist, B. H., 2016. Nonhomogeneous poisson process with nonparametric frailty. Reliability Engineering & System Safety 149, 14–23.

[15] Slimacek, V., Lindqvist, B. H., 2016. Reliability of wind turbines modeled by a poisson process with covariates, unobserved heterogeneity and seasonality. Wind Energy 19 (11), 1991–2002, we.1964.
URL `http://dx.doi.org/10.1002/we.1964`

[16] Tavner, P., Greenwood, D., Whittle, M., Gindele, R., Faulstich, S., Hahn, B., 2013. Study of weather and location effects on wind turbine failure rates. Wind Energy 16 (2), 175–187.

[17] Tavner, P., Xiang, J., Spinato, F., 2007. Reliability analysis for wind turbines. Wind Energy 10 (1), 1–18.

[18] Van Dyck, J., Verdonck, T., 2014. Precision of power-law nhpp estimates for multiple systems with known failure rate scaling. Reliability Engineering & System Safety 126, 143–152.

[19] Wu, S., Scarf, P., 2015. Decline and repair, and covariate effects. European Journal of Operational Research 244 (1), 219–226.

[20] You, M., Meng, G., 2012. Residual life prediction of repairable systems subject to imperfect preventive maintenance using extended propor-

tional hazards model. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering 226 (1), 50–63.

Appendix A. Variance of the estimator $\hat{\boldsymbol{\beta}}$

Recall that the measured covariates $\boldsymbol{x}_j$'s and the observation times τ_j's ($j = 1, \ldots, m$) are observed parameters describing the process, since they are known before the start of the process and they do not change during the process. The sample version of the function $h(\boldsymbol{\beta})$ will therefore be viewed as a function of the observed number of events and the estimator $\hat{b}$ (which together form an observed data), i.e. the notation $\hat{h}(\boldsymbol{\beta}, \boldsymbol{d})$, where $\boldsymbol{d} = [n_1, \ldots, n_m, \hat{b}]$, will be used. All the computations below are done conditionally on $\boldsymbol{x}_j$'s and τ_j's ($j = 1, \ldots, m$), which will be, in most of cases, omitted from the notation. The set of all $\boldsymbol{x}_j$'s, $j = 1, \ldots, m$, will be denoted by $\boldsymbol{x}$ and the set of all τ_j's, $j = 1, \ldots, m$, will be denoted by $\boldsymbol{\tau}$.

The minimum of the function $\hat{h}(\boldsymbol{\beta}, \boldsymbol{d})$ is described by the set of equations

$$\boldsymbol{\Phi}(\boldsymbol{\beta}, \boldsymbol{d}) = \left\{ \Phi_r(\boldsymbol{\beta}, \boldsymbol{d}) = \frac{\partial \hat{h}(\boldsymbol{\beta}, \boldsymbol{d})}{\partial \beta_r},\ r = 1, \ldots, R \right\} = \boldsymbol{0} \tag{A.1}$$

which has a solution $\boldsymbol{\beta} = \frac{\hat{\boldsymbol{\beta}}}{2}$ as was shown above.

Let us assume, that

$$\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{\beta}} \left(\frac{\hat{\boldsymbol{\beta}}}{2}, \boldsymbol{d} \right) = \begin{bmatrix} \frac{\partial \Phi_1}{\partial \beta_1} & \cdots & \frac{\partial \Phi_1}{\partial \beta_R} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Phi_R}{\partial \beta_1} & \cdots & \frac{\partial \Phi_R}{\partial \beta_R} \end{bmatrix} \left(\frac{\hat{\boldsymbol{\beta}}}{2}, \boldsymbol{d} \right) \tag{A.2}$$

is an invertible matrix. This assumption is not very restrictive, since the sample version of the function $h(\boldsymbol{\beta})$ is strictly convex. The implicit function theorem then ensures existence of a function

$$\boldsymbol{\varphi}(\boldsymbol{d}) = \{\varphi_r(\boldsymbol{d}), r = 1, \ldots, R\} = \frac{\hat{\boldsymbol{\beta}}}{2} \tag{A.3}$$

which maps the observed data $\boldsymbol{d} = [n_1, \ldots, n_m, \hat{b}]$ into an estimate $\frac{\hat{\boldsymbol{\beta}}}{2}$.

As can be easily checked by minimization of the function $\hat{h}(\boldsymbol{\beta}, \mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}))$, where $\mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ represents the conditionally expected data, (which were computed with use of the individuals version of (7), properties of the estimator $\hat{b}$ and the double expectation rule),

$$\begin{aligned} \mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}) &= \left[\mathrm{E}(n_1|\boldsymbol{x}_1, \tau_1), \ldots, \mathrm{E}(n_m|\boldsymbol{x}_m, \tau_m), \mathrm{E}\left(\hat{b}|\boldsymbol{x}, \boldsymbol{\tau}\right) \right]' \\ &= \left[a\tau_1^b \exp\left(\boldsymbol{\beta}_0' \boldsymbol{x}_1\right), \ldots, a\tau_m^b \exp\left(\boldsymbol{\beta}_0' \boldsymbol{x}_m\right), b \right]' \end{aligned} \tag{A.4}$$

this function maps the conditionally expected data into the point $\frac{\boldsymbol{\beta_0}}{2}$, i.e. $\boldsymbol{\varphi}(\mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})) = \frac{\boldsymbol{\beta_0}}{2}$. Taylor expansion of the function $\boldsymbol{\varphi}(\boldsymbol{d})$ around the conditionally expected data $\mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ therefore gives

$$\frac{\hat{\boldsymbol{\beta}}}{2} = \boldsymbol{\varphi}(\boldsymbol{d}) \approx \frac{\boldsymbol{\beta_0}}{2} + \nabla \boldsymbol{\varphi}(\mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}))' \left[\boldsymbol{d} - \mathrm{E}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}) \right] \tag{A.5}$$

where $\nabla = \left[\frac{\partial}{\partial n_1}, \ldots, \frac{\partial}{\partial n_m}, \frac{\partial}{\partial \hat{b}} \right]$ denotes the (row) gradient operator.

The equation (A.5) can also be used for the computation of the conditional covariance matrix of $\frac{\hat{\beta}}{2}$, which gives the well-known approximation

$$\operatorname{Cov}\left(\frac{\hat{\boldsymbol{\beta}}}{2}\middle|\, \boldsymbol{x}, \boldsymbol{\tau}\right) \approx \nabla\boldsymbol{\varphi}\left(\mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right)' \operatorname{Cov}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right) \nabla\boldsymbol{\varphi}\left(\mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) \tag{A.6}$$

where $\operatorname{Cov}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ is the covariance matrix of the observed numbers of events n_j's and the estimator $\hat{b}$ given the measured covariates $\boldsymbol{x}_j$'s and the observation times τ_j's ($j = 1, \ldots, m$). Since the systems are assumed to be independent and $\operatorname{Cov}\left(n_j, \hat{b}\middle|\, \boldsymbol{x}, \boldsymbol{\tau}\right) = 0$ for each $j = 1, \ldots, m$ (which follows from the properties of the estimator $\hat{b}$ and the conditional covariance rule), the covariance matrix $\operatorname{Cov}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ is a diagonal matrix where the first m diagonal terms can be computed as (by using the individuals versions of (7) and (8))

$$\begin{aligned}\operatorname{Var}\left(N_j|\boldsymbol{X}_j, \tau_j\right) &= \mathrm{E}\left(\operatorname{Var}\left(N_j|Z_j, \boldsymbol{X}_j, \tau_j\right)\right) + \operatorname{Var}\left(\mathrm{E}\left(N_j|Z_j, \boldsymbol{X}_j, \tau_j\right)\right) \\ &= a\tau_j^b \exp\left(\boldsymbol{\beta}_0' \boldsymbol{X}_j\right) + \left(a\tau_j^b \exp\left(\boldsymbol{\beta}_0' \boldsymbol{X}_j\right)\right)^2 \operatorname{Var}(Z)\end{aligned} \tag{A.7}$$

where $\operatorname{Var}(Z)$ denotes the variance of the unobserved effects. The $(m+1)$st term on diagonal of the matrix $\operatorname{Cov}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ is equal to $\operatorname{Var}\left(\hat{b}\middle|\, \boldsymbol{x}, \boldsymbol{\tau}\right)$.

If it would be possible to find the analytical expression of the function (A.3), the equation (A.6) together with $\operatorname{Cov}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ would define the conditional covariance matrix of the estimator $\hat{\boldsymbol{\beta}}$. Although the function (A.3) can not be found analytically, the implicit function theorem gives the formula for the desired gradient of this function, i.e.,

$$\nabla\boldsymbol{\varphi}\left(\mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) = -\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{\beta}}\left(\frac{\boldsymbol{\beta}_0}{2}, \mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right)^{-1} \frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{d}}\left(\frac{\boldsymbol{\beta}_0}{2}, \mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) \tag{A.8}$$

where $\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{\beta}}\left(\frac{\boldsymbol{\beta}_0}{2}, \mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right)$ is defined by (A.2) and $\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{d}}\left(\frac{\boldsymbol{\beta}_0}{2}, \mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right)$ is analogously defined as

$$\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{d}}\left(\frac{\boldsymbol{\beta}_0}{2}, \mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) = \begin{bmatrix} \frac{\partial \Phi_1}{\partial n_1} & \cdots & \frac{\partial \Phi_1}{\partial n_m} & \frac{\partial \Phi_1}{\partial \hat{b}} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \Phi_R}{\partial n_1} & \cdots & \frac{\partial \Phi_R}{\partial n_m} & \frac{\partial \Phi_R}{\partial \hat{b}} \end{bmatrix} \left(\frac{\boldsymbol{\beta}_0}{2}, \mathrm{E}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) \tag{A.9}$$

Using the estimators of in the derived formulas gives the estimator of the conditional covariance of the estimator $\hat{\boldsymbol{\beta}}$, i.e.,

$$\widehat{\operatorname{Cov}}\left(\hat{\boldsymbol{\beta}}\middle|\, \boldsymbol{x}, \boldsymbol{\tau}\right) = 4\nabla\boldsymbol{\varphi}\left(\widehat{\mathrm{E}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right)' \widehat{\operatorname{Cov}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right) \nabla\boldsymbol{\varphi}\left(\widehat{\mathrm{E}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) \tag{A.10}$$

where

$$\nabla\boldsymbol{\varphi}\left(\widehat{\mathrm{E}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) = -\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{\beta}}\left(\frac{\hat{\boldsymbol{\beta}}}{2}, \widehat{\mathrm{E}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right)^{-1} \frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{d}}\left(\frac{\hat{\boldsymbol{\beta}}}{2}, \widehat{\mathrm{E}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right)\right) \tag{A.11}$$

The estimated conditional expected value $\widehat{\mathrm{E}}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ is defined with use of (A.4), i.e.,

$$\widehat{\mathrm{E}}\left(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau}\right) = \left[\hat{a}\tau_1^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_1\right), \ldots, \hat{a}\tau_m^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_m\right), \hat{b}\right]' \tag{A.12}$$

The estimated conditional covariance matrix $\widehat{\operatorname{Cov}}(\boldsymbol{d}|\boldsymbol{x}, \boldsymbol{\tau})$ is a diagonal matrix with the jth entry (for $j = 1, \ldots, m$) on the diagonal equal to $\hat{a}\tau_j^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j\right) + \left(\hat{a}\tau_j^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j\right)\right)^2 \widehat{\operatorname{Var}}(Z)$ (estimated with use of (A.7)), where $\widehat{\operatorname{Var}}(Z)$ is the estimator of the variance of the unobserved effects derived below, and the $(m+1)$st diagonal element defined as $\frac{\hat{b}^2}{\sum_{j=1}^{m} n_j}$ (defined with use $\widehat{\text{Var}}\left(\hat{b}\right)$ and the conditional variance formula).

The function $\mathbf{\Phi}(\boldsymbol{\beta}, \boldsymbol{d})$ and its partial derivatives are defined by (A.1), (A.2) and (A.9) respectively.

Note, that the analytical expressions for the function and $\mathbf{\Phi}(\boldsymbol{\beta}, \boldsymbol{d})$ and its partial derivatives can easily be found from the sample version of the function $h(\boldsymbol{\beta}, \boldsymbol{d})$.

Appendix B. Variance of the estimator $\hat{a}$

The variance of $\hat{a}$ (given $\boldsymbol{x}_j$'s and τ_j's , $j = 1, \ldots, m$) can be derived with use of Taylor expansion and properties of the estimators $\hat{b}$ and $\hat{\boldsymbol{\beta}}$ and of NHPPs. Taking the Taylor expansion of the expression $\frac{n_j}{\tau_j^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j\right)}$ around the points $\mathrm{E}\left(\hat{b}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right) = b$, $\mathrm{E}\left(\hat{\boldsymbol{\beta}}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right) = \boldsymbol{\beta}_0$ and $\mathrm{E}\left(n_j | \boldsymbol{x}, \boldsymbol{\tau}\right) = a\tau_j^b \exp\left(\boldsymbol{\beta}_0' \boldsymbol{x}_j\right)$ gives

$$\frac{n_j}{\tau_j^{\hat{b}} \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j\right)} \approx \frac{n_j}{\tau_j^{b} \exp\left(\boldsymbol{\beta}_0' \boldsymbol{x}_j\right)} - a \log(\tau_j)(\hat{b} - b) - a\boldsymbol{x}_j'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \tag{B.1}$$

Taking the conditional variance of $\hat{a}$ with use of (B.1) and properties of the estimators $\hat{b}$ and $\hat{\boldsymbol{\beta}}$ and nonhomogeneous Poisson processes then gives

$$\begin{aligned}
\mathrm{Var}\left(\hat{a}|\boldsymbol{x}, \boldsymbol{\tau}\right) \approx {} & \frac{a}{m^2} \sum_{j=1}^{m} \frac{1}{\tau_j^b \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j\right)} + \frac{a^2}{m} \mathrm{Var}\left(Z\right) \\
& + \left(\frac{a}{m} \sum_{j=1}^{m} \log(\tau_j)\right)^2 \mathrm{Var}\left(\hat{b}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right) + \left(\frac{a}{m} \sum_{j=1}^{m} \boldsymbol{x}_j\right)' \mathrm{Cov}\left(\hat{\boldsymbol{\beta}}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right) \left(\frac{a}{m} \sum_{j=1}^{m} \boldsymbol{x}_j\right) \\
& + 2\left(\frac{a}{m} \sum_{j=1}^{m} \log(\tau_j)\right) \left(\frac{a}{m} \sum_{j=1}^{m} \boldsymbol{x}_j\right)' \mathrm{Cov}\left(\hat{b}, \hat{\boldsymbol{\beta}}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right) \\
& - 2\left(\frac{a}{m} \sum_{j=1}^{m} \boldsymbol{x}_j\right)' \left(\frac{1}{m} \sum_{j=1}^{m} \frac{\mathrm{Cov}\left(n_j, \hat{\boldsymbol{\beta}}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right)}{\tau_j^b \exp\left(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_j\right)}\right)
\end{aligned} \tag{B.2}$$

Substitution of unknown parameters by their estimators in (B.2) defines the estimator of the conditional variance of the estimator $\hat{a}$. The unknown covariances $\mathrm{Cov}\left(\hat{b}, \hat{\boldsymbol{\beta}}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right)$ and $\mathrm{Cov}\left(n_j, \hat{\boldsymbol{\beta}}\middle| \boldsymbol{x}, \boldsymbol{\tau}\right)$, $j = 1, \ldots, m$ can be estimated with use of (A.5) and properties of the estimator $\hat{b}$ and of NHPPs.