WILEY

Prediction intervals for load-sharing systems in accelerated life testing

Kevin Leckey[1] | Christine H. Müller[1] | Sebastian Szugat[1] | Reinhard Maurer[2]

[1]Faculty of Statistics, TU Dortmund University, Dortmund, Germany

[2]Faculty of Architecture and Civil Engineering, TU Dortmund University, Dortmund, Germany

Correspondence
Kevin Leckey, TU Dortmund University, D-44227 Dortmund, Germany.
Email: kevin.leckey@tu-dortmund.de

Abstract

Based on accelerated lifetime experiments, we consider the problem of constructing prediction intervals for the time point at which a given number of components of a load-sharing system fails. Our research is motivated by lab experiments with prestressed concrete beams where the tension wires fail successively. Due to an audible noise when breaking, the time points of failure could be determined exactly by acoustic measurements. Under the assumption of equal load sharing between the tension wires, we present a model for the failure times based on a birth process. We provide a model check based on a Q-Q plot including a simulated simultaneous confidence band and four simulation-free prediction methods. Three of the prediction methods are given by confidence sets where two of them are based on classical tests and the third is based on a new outlier-robust test using sign depth. The fourth method uses the implicit function theorem and the δ-method to get prediction intervals without confidence sets for the unknown parameter. We compare these methods by a leave-one-out analysis of the data on prestressed concrete beams. Moreover, a simulation study is performed to discuss advantages and drawbacks of the individual methods.

KEYWORDS

reliability, survival analysis, statistical quality control

1 | INTRODUCTION

We consider J load-sharing systems that are exposed to different initial stress conditions $s_1, \ldots, s_J$. Each system consists of I components, and the external load of the system is equally distributed over all working components. In particular, the stress of a working component increases if other components have failed. This is also known as load redistribution with equal load-share rule. A system fails if a critical number $I_c \leq I$ of its components fails. We assume that the time points of component failures are observed up to $I_j \leq I$ for $j = 1, \ldots, J$. Our aim is to provide prediction intervals for the time point that I_c components have failed in a new system exposed to a specific initial stress s_0. The initial stress of the new system, for which the prediction should be done, can be much lower than the initial stress of observed systems, that is, $s_0 < \min\{s_1, \ldots, s_J\}$. Hence, we consider the problem of accelerated lifetime experiments.

Our research is motivated by experiments with prestressed concrete beams with $I = 35$ tension wires that are exposed to cyclic loading. In this case, the stress on the beam is periodically increased/decreased between two fixed values. The difference between these two values is called stress range and takes the role of the load/stress s_j in our model. Other examples[1] of equal load-sharing system include systems of electrical components or batteries in a parallel arrangement.

Our experimental setting is different from the situation of systems of I independent components often considered in reliability analysis and usually called k-out-of-n systems with $k = I_c$ and $n = I$ in our notation. In such situations, the life-

times of the components can be assumed to be independent and identically distributed. For several lifetime distributions, prediction intervals for the time of the I_cth failure are constructed if $I_0 < I_c \leq I$ failures of I components were observed. This is usually done by predicting the I_cth-order statistic based on the I_0th-order statistic.[2-4] Nordman and Meeker[5] also construct a prediction interval for the additional number of failures in the I components at time t_w based on the number of failures at time $t < t_w$.

According to our knowledge, prediction intervals for load-sharing systems have not been derived up to now. Even load-sharing systems exposed to different initial load conditions are not treated, although several results for identically distributed load-sharing systems have been published recently. Most of these results are based on sequential order statistics.[6-11] Several approaches assume exponential or Weibull distribution[12-17] or other parametric distributions[18] for the lifetime of the components or the time between successive failures. Semiparametric approaches are given by Kvam and Peña[1] and Deshpande et al.[19] Kong and Ye use a cumulative exposure model[20] as it was used in step-stress accelerated life testing and provide a method[21] for testing the exponential distribution for the time between successive failures. Confidence sets for model parameters were derived for some special parametric models.[12,15,22] Xu et al[16] propose a test for the mean time to failure in a load-sharing model with Weibull distribution.

In most of these approaches, I or I_c parameters must be estimated. This is only possible if the number I of components and the number I_c of critical failures is not too large and several independent and identical systems under same conditions can be observed. However, here we will consider systems with more than 30 components, and the systems are observed under different stress conditions. To get reliable prediction intervals for the time that a critical number I_c of components has failed, it is important to use a model with not too many unknown parameters to avoid overfitting.

There are some simple models with few parameters for identically distributed load-sharing systems based on Weibull distribution[13] and exponential distribution.[15] To reduce complexity, we assume an exponential distribution for the times between failures, but with the difference that we incorporate the different initial stress conditions by a link function. This leads to nonlinear birth processes.

We present different methods for simulation-free prediction intervals for the time until the I_cth failure (event) in a new system under initial stress s_0 when other systems exposed to different initial stress levels have been observed before. We also include the possibility that some failures of the new system have already been detected.

Our approach is based on the fact that exponentially distributed interarrival times lead to failure times with a hypoexponential distribution and in particular the distribution function is known explicitly. However, a standard implementation of this distribution function can become numerically instable, which can cause the implemented function to not be monotone increasing and to take values above 1. Fortunately, there is an efficient and more stable algorithm for calculating the hypoexponential distribution function.[23] Hence, the quantiles of the failure time distribution can be computed efficiently and only the uncertainty of the unknown model parameters needs to be incorporated.

A simple possibility to include this uncertainty is to use classical confidence sets for the parameters of generalized linear models based on, for example, the Wald or the likelihood ratio test.[24, pp. 409] Here, we propose two alternatives. One alternative is based on the sign depth for getting outlier-robust confidence sets.[25,26] Since all methods based on confidence sets for the parameter vector have the shortcoming that they require a grid search or another form of candidate generation to evaluate the respective test, we also use the δ-method to construct confidence intervals for the quantiles of the predictive distribution as a second alternative.

The paper is organized as follows. Section 2 introduces the model including the prediction problem and provides a model check based on a Q-Q plot with a simulated simultaneous confidence band. Section 3 deals with the derivation of confidence sets. Section 3.1 summarizes classical confidence sets for the whole vector of model parameters and one-dimensional aspects of the model parameters such as the quantiles of the predictive distribution. In particular, the δ-method for asymptotic distributions is used to obtain confidence intervals for one-dimensional aspects of the model parameters. Completely new outlier-robust confidence sets for the whole model parameter are derived in Section 3.2. Section 4 contains the derivation of prediction intervals starting with the predictive distribution in Section 4.1. Numerical problems when calculation the quantiles of this so-called hypoexponential distribution are also discussed here. Prediction intervals based on confidence sets for the whole model parameter are given in Section 4.3 whereas predictions based on confidence intervals for quantiles of the predictive distribution can be found in in Section 4.4. In Section 5, we apply the model check and the different prediction methods to the data of experiments with prestressed concrete beams and compare the performance of the prediction methods by cross-validation. A comparison via simulation is given in Section 6. We summarize and discuss our results in Section 7. Appendix A.1 provides the mathematical foundations for getting the classical confidence sets, Appendix A.2 for all proofs, and Appendix A.3 details about the R implementation.

2 | THE LOAD-SHARING MODEL AND THE PREDICTION PROBLEM

In this section, we provide the load-sharing model (Section 2.1), the prediction problem in the load-sharing model (Section 2.2), and a model check of the load-sharing model (Section 2.3).

2.1 | The load-sharing model

We assume that each system $j = 0, 1, \ldots, J$ has I components and that the failure times of the components of system j are given by a point process $0 = T_{0,j} < T_{1,j} < T_{2,j} < \ldots < T_{I_j,j}$ with $I_j \leq I$, where I_j is the number of observed failures in the jth system. It can be assumed that all these time points are different although, in reality, several failures may be observed at the same time since the time is measured on a discrete scale. Let $W_{i,j} = T_{i,j} - T_{i-1,j}$ and $i \in \mathbb{N}$ be the waiting times (interarrival times) between the $(i-1)$th and ith failures (events) and let

$$N^j(t) := N_t^j := \sum_{i=1}^{\infty} \mathbb{1}_{[0,t]}(T_{i,j}), \ t \geq 0$$

be the corresponding counting process of the failures, where $\mathbb{1}_{[0,t]}(x)$ denotes the indicator function which equals one if $x \in [0, t]$.

Here, we assume that only the number of past failures (events) $N_{t^-}^j = \lim_{\tau \uparrow t} N_\tau^j$ influences the intensity of appearance of a failure at time t. Hence, $(N_t^j)_{t \geq 0}$ is a state dependent point process, that is, a birth process.[27, p. 95, 96][28, p. 21, p. 95] In particular, we regard nonlinear birth processes given by the intensity function

$$\lambda^j(t) := h^j \left(\frac{I}{I - N_{t^-}^j} \right).$$

This implies that the waiting times $W_{i+1,j}$ and $i = 0, 1, \ldots, I_j - 1$ are independent and have an exponential distribution with parameter $\lambda_i^j := h^j (I/(I-i))$. Since we aim to model several independent systems under different initial stress levels, we incorporate the initial stress level s_j by

$$\lambda_i^j = \lambda_\theta(i, s_j) := h_\theta \left(s_j \cdot \frac{I}{I - i} \right),$$

where h_θ is a given link function between stress and failure time and $\theta \in \Theta \subset \mathbb{R}^p$ is an unknown parameter. Hence, the waiting times are independent and exponentially distributed with

$$W_{i,j} \sim \text{Exp}(\lambda_\theta(i-1, s_j)), \ i = 1, \ldots, I_j \leq I, \ j = 0, 1, \ldots, J.$$

An example for the function h_θ is

$$h_\theta(x) := \exp(-\theta_1 + \theta_2 \ln(x)) \tag{1}$$

with $\theta = (\theta_1, \theta_2)^\top \in \Theta = \mathbb{R} \times [0, \infty)$ so that the logarithm of the expected waiting time

$$\log(\text{E}(W_{i,j})) = \log \left(\frac{1}{\lambda_\theta(i-1, s_j)} \right) = \theta_1 - \theta_2 \ln \left(s_j \cdot \frac{I}{I - (i-1)} \right)$$

is strictly decreasing in s_j and i. It is the often used linear link function proposed by Basquin.[29] However, other linear and nonlinear link functions are possible.

Our model is different from the linear birth process given by intensity $\lambda_i = i\lambda$ for fatigue accumulation mentioned by Sobczyk and Spencer.[30, p.112] However, for load redistribution, our model is very reasonable since the failure of αI components increases the individual load on each of the remaining $(1-\alpha)I$ components by a factor of $1/(1-\alpha)$. Hence, the load increases nonlinearly. Note that this load-sharing rule has previously been used[12,31] but without the function h. The Basquin link was also used by Kong and Ye[15] for load-sharing models based on exponentially distributed lifetimes of the components. We are aware that our model does not include damage accumulation and aging. This is only possible by more general self-exciting point processes that can model damage accumulation and aging via intensity functions that are not constant between failures, see Section 7. However, this makes the derivation of simulation-free prediction intervals difficult. In particular, only exponential waiting times yield a Markov process.[32, Theorem 8.2.9] Moreover, we have fairly small

datasets in mind (see Section 5), which always bears the risk of overfitting if models with too many model parameters are used. Hence, we keep the model as simple as possible. However, we provide a model check in Section 2.3. The results for the real data set in Section 5 indicate that this model is realistic enough to provide good prediction intervals.

2.2 | The prediction problem

The aim of this article is to provide prediction intervals for the time at which a critical number I_c of the I components has failed. Depending on the application, I_c could be I but a smaller value is possible as well, for example, if the stability of the system cannot be guaranteed after a certain number of component failures. Moreover, the number I_j of observed component failures in experiment j is not restricted to be exactly I_c. A smaller value $I_j < I_c$ usually appears if the corresponding experiment lasts too long and thus is stopped due to time/cost limitations. Larger values $I_j > I_c$ are possible if I_c is only the minimum number of failures for which some safety condition of the system is not ensured.

The predictions are based on observations from previous experiments as well as potentially a certain number of component failures in the current system for which a prediction is made. More precisely, we observe realizations $w_{i,j}$ of the random variables $W_{i,j}$, $i = 1, \ldots, I_j$, and $j = 1, \ldots, J$, from J different systems and eventually additional realizations $w_{i,0}$ of the random variables $W_{i,0}$, $i = 1, \ldots, I_0$, from a new system. For the new system, I_0 is smaller than I_c and is set equal to 0 if no failure of the new process was observed. In particular, $w_{1,0}, \ldots, w_{I_0,0}$ are already realizations of the random variables $W_{1,0}, \ldots, W_{I_0,0}$ while the random variables $W_{I_0+1,0}, \ldots, W_{I_c,0}$ will be realized in future. The aim is to predict

$$w_{1,0} + \ldots + w_{I_0,0} + W_{I_0+1,0} + \ldots + W_{I_c,0}, I_c > I_0, \tag{2}$$

that is, the time up to the failure of the I_cth component of the new system. Note that it is sufficient to predict $W_{I_0+1,0} + \ldots + W_{I_c,0}$ since this immediately* yields a prediction interval for (2). To quantify the uncertainty of a point prediction, we derive prediction intervals in Section 4 based on confidence sets.

2.3 | A model check

Although there are many tests for checking the exponential distribution for i.i.d. variables,[33] the problem of checking the exponential distribution for variables that are not identically distributed is challenging. Kong and Ye[21] provides a goodness-of-fit test for their approach of a load sharing model with exponentially distributed lifetimes. Since any goodness-of-fit test has the disadvantage of a 0-1 decision of rejecting or not rejecting the null hypotheses, we present here a model check via a Q-Q plot including a simulated simultaneous 90%-confidence band.

Since the waiting times are not identically distributed, this requires some kind of standardization first: The sample quantiles in the Q-Q plot are taken from the rescaled waiting times

$$\tilde{W}_{i,j} = \lambda_{\hat{\theta}}(i - 1, s_j) \cdot W_{i,j}, \tag{3}$$

where $\hat{\theta}$ is the maximum likelihood estimator for θ and the rates are chosen according to the Basquin model, that is,

$$\lambda_\theta(i - 1, s_j) = e^{-\theta_1} \cdot \left(\frac{s_j I}{I - (i - 1)} \right)^{\theta_2} \quad \text{for} \quad \theta = (\theta_1, \theta_2) \in \mathbb{R} \times [0, \infty). \tag{4}$$

Note that if $W_{i,j}$ is distributed according to our model and the true model parameter is θ, then $\lambda_\theta(i-1, s_j) \cdot W_{i,j} \sim \text{Exp}(1)$. Hence, $\tilde{W}_{i,j}$ should have a distribution close to Exp(1) if $\hat{\theta}$ is a consistent estimator for θ.

We will discuss our approach for deriving a simultaneous confidence band in the following more general framework. Consider independent random variables $W_1(\theta), \ldots, W_N(\theta)$ from the parametric model

$$W_n(\theta) \sim \text{Exp}(\lambda_n(\theta)), \quad n = 1, \ldots, N,$$

---

*the condition $W_{1,0} + \ldots + W_{I_0,0} = w_{1,0} + \ldots + w_{I_0,0}$ can be disregarded since $W_{i,0}$, $i = 1, \ldots, I$, are independent by assumption

with $\theta \in \Theta \subset \mathbb{R}^p$, $p \geq 1$ and $\lambda_n : \Theta \to (0, \infty)$, where N is the number of observations (here $N := N(J) := \sum_{j=0}^{J} I_j$).

Suppose that the rates satisfy

$$\frac{\lambda_n(\theta)}{\lambda_n(\tilde{\theta})} = h_n(\theta - \tilde{\theta}), \quad n = 1, \dots, N \tag{5}$$

for all $\theta, \tilde{\theta} \in \Theta$ and some function $h_n : \mathbb{R}^p \to \mathbb{R}$. Note that condition (5) is satisfied for the Basquin model given by (4). We first observe that a rescaling with the maximum likelihood estimation (MLE) of θ leads to a distribution that does not depend on the model parameter anymore, hence allowing us to fix θ to any value in the subsequent analysis. Note that this was already observed for i.i.d. exponentially distributed random variables,[34] in which case the MLE is given by the inverse of the sample mean.

Lemma 1. *Let $\hat{\theta}(\theta)$ denote the MLE for θ based on the data $(W_1(\theta), \dots, W_N(\theta))$ and*

$$\tilde{W}_n(\theta) := \lambda_n(\hat{\theta}(\theta)) \cdot W_n(\theta), \quad n = 1, \dots, N.$$

If $\lambda_n(\theta)$ satisfies (5), then the distribution of $(\tilde{W}_1(\theta), \dots, \tilde{W}_N(\theta))$ does not depend on θ, that is,

$$(\tilde{W}_1(\theta), \dots, \tilde{W}_N(\theta)) \stackrel{d}{=} (\tilde{W}_1(\tilde{\theta}), \dots, \tilde{W}_N(\tilde{\theta})) \quad \text{for all} \quad \theta, \tilde{\theta} \in \Theta.$$

Computing the confidence band. The confidence band is based on an adaptation of the Kolmogorov-Smirnov test to the rescaled random variables in (3). To this end, fix an arbitrary parameter θ (the chosen parameter is irrelevant according to Lemma 1). Let M be the number of simulations used to determine the confidence band, that is, consider random variables

$$X_{n,m} := \tilde{W}_{n,m}(\theta) := \lambda_n(\hat{\theta}_m(\theta)) \cdot W_{n,m}(\theta), \quad n = 1, \dots, N, \ m = 1, \dots, M,$$

where $(W_{1,m}(\theta), \dots, W_{N,m}(\theta))$, $m = 1, \dots, M$, are independent copies of $(W_1(\theta), \dots, W_N(\theta))$ and $\hat{\theta}_m(\theta)$ is the ML estimate for θ based on $W_{1,m}(\theta), \dots, W_{N,m}(\theta)$. Let $x_{1,m}, \dots, x_{N,m}$ be realizations of $X_{1,m}, \dots, X_{N,m}$ and let $x_{(1),m}, \dots, x_{(N),m}$ denote their order statistic for $m = 1, \dots, M$. Furthermore, let F_m denote the empirical distribution function of $x_{1,m}, \dots, x_{N,m}$, that is,

$$F_m(x) := \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[0,x]}(x_{n,m}), \quad x \geq 0, \quad m = 1, \dots, M.$$

Let $F(x) = 1 - e^{-x}$, $x \geq 0$ denote the distribution function of $\mathrm{Exp}(1)$ and q_n the $\frac{n}{N+1}$-quantile of this distribution. Let $\|f\|_\infty = \sup_{t\in[0,\infty)} |f(t)|$ denote the sup norm of a function $f : [0, \infty) \to \mathbb{R}$. Note that since F is increasing and F_m is piecewise constant, $x \mapsto F_m(x) - F(x)$ is a monotone decreasing function between two consecutive jumps of F_m (note that this difference might change from positive to negative and thus the absolute difference $x \mapsto |F_m(x) - F(x)|$ is not a monotone function). Hence, the absolute value of this difference attains its global maximum at the beginning or end of a jump, that is, if $a < b$ are two consecutive jumps of F_m, then

$$\sup_{x\in[a,b)} |F_m(x) - F(x)| = \max \left\{ \lim_{x\uparrow b} |F_m(x) - F(x)|, \ |F_m(a) - F(a)| \right\},$$

where $\lim_{x\uparrow b}$ denotes the limit to b from below. Since jumps of F_m occur at $x_{(1),m}, \dots, x_{(N),m}$ and since $F_m(x_{(n),m}) = n/N$ and $\lim_{x\uparrow x_{(n),m}} F_m(x) = (n-1)/N$, we therefore obtain

$$\|F_m - F\|_\infty = \max \left\{ \limsup_{x\to x_{n,m}} |F_m(x) - F(x)|; \ n = 1, \dots, N \right\}$$

$$= \max \left\{ \max \left\{ \left| \frac{n-1}{N} - F(x_{(n),m}) \right|, \ \left| \frac{n}{N} - F(x_{(n),m}) \right| \right\}; \ n = 1, \dots, N \right\}.$$

Finally, let $c_{0.90}$ be the empirical 90%-quantile of $\|F_1 - F\|_\infty, \ldots, \|F_M - F\|_\infty$. Then, the simulated simultaneous 90%-confidence band for the Q-Q plot is the smallest band containing all the points in

$$\left\{ (q_1, x_{(1),m}), \ldots, (q_N, x_{(N),m}); \; m \quad \text{with} \quad \|F_m - F\|_\infty \leq c_{0.90} \right\}.$$

More formally, the upper bound of the confidence band is a linear interpolation of the points $(q_1, \max_{m \in A} x_{(1),m}), \ldots, (q_N, \max_{m \in A} x_{(N),m})$ where $A = \{m; \|F_m - F\|_\infty \leq c_{0.90}\}$. The lower bound of the band is obtained by taking the minimum instead.

> *Remark* 1. Note that $X_{1,m}, \ldots, X_{N,m}$ are not identically distributed (their distributions only converge to Exp(1)). We therefore determine the quantile $c_{0.90}$ by simulation rather than taking the critical value of the Kolmogorov-Smirnov test for the Exp(1) distribution.

3 | CONFIDENCE SETS

Most of the prediction intervals in Section 4 are based on confidence sets for θ. We will therefore discuss two classical approaches as well as a new outlier-robust approach to derive such confidence sets in this section. Another prediction approach in Section 4.4 is based on confidence intervals for the quantiles of the predictive distribution. Hence, we will also provide a confidence interval for a one-dimensional aspect $g(\theta)$ of the parameter. As an overview, the following four confidence sets/intervals are provided:

(a) A classical confidence set for θ based on the Wald test (6) in Section 3.1.
(b) A classical confidence set for θ based on the likelihood-ratio test (8) in Section 3.1.
(c) A confidence interval (9) for a one-dimensional aspect $g(\theta)$, $g : \Theta \to \mathbb{R}$, in Section 3.1.
(d) A new outlier-robust confidence set for θ based on sign depth (10) in Section 3.2.

Note that the sets in (a) to (c) are affected quite a lot by outliers in the data, that is, by observations that are atypically small or large. Method (d) provides an alternative that is less sensitive to such outliers but has the disadvantage of yielding a larger confidence set, in particular for small sample sizes. Hence, an appropriate confidence set from (a) to (d) should be chosen based on whether outliers are likely to occur in the application.

3.1 | Classical confidence sets for the whole parameter and for functions of the parameter

Since we assume that the waiting times (interarrival times) $W_{i,j}$ are independent for $i = 1, \ldots, I_j$ and $j = 0, 1, \ldots, J$ and have an exponential distribution with parameter $\lambda_\theta(i - 1, s_j)$, we can easily derive a maximum likelihood estimator $\widehat{\theta}$ for θ by

$$\widehat{\theta} = \arg \max_\theta \prod_{j=0}^{J} \prod_{i=1}^{I_j} f_{\lambda_\theta(i-1,s_j)}(w_{i,j}),$$

where $f_\lambda(w)$ denotes the density of the exponential distribution with parameter λ and $w_{i,j}$ is the realization of $W_{i,j}$ for all i, j. A classical $(1 - \alpha)$-confidence set for the whole parameter vector θ based on this maximum likelihood estimator $\widehat{\theta}$ is given by

$$\left\{ \theta; \; (\widehat{\theta} - \theta)^\top I(\widehat{\theta})(\widehat{\theta} - \theta) \leq \chi^2_{p,1-\alpha} \right\}, \tag{6}$$

where

$$I(\theta) := \sum_{j=0}^{J} \sum_{i=1}^{I_j} \frac{1}{\lambda_\theta(i-1,s_j)^2} \dot{\lambda}_\theta(i-1,s_j) \, \dot{\lambda}_\theta(i-1,s_j)^\top \quad \text{with} \quad \dot{\lambda}_\theta(i,s) := \frac{\partial}{\partial \theta} \lambda_\theta(i,s) \tag{7}$$

is the information matrix, $\chi^2_{p,1-\alpha}$ denotes the $(1 - \alpha)$-quantile of the χ^2-distribution with p degrees of freedom and p is the dimension of the model parameter, that is, $\theta \in \mathbb{R}^p$. The related test is widely known as the Wald test.[37, p. 349] Another

classical $(1 - \alpha)$-confidence set for θ based on the likelihood-ratio test[35, pp. 459-461] is given by

$$
\left\{ \theta; -2 \ln \left(\frac{\prod\limits_{j=0}^{J} \prod\limits_{i=1}^{I_j} f_{\lambda_\theta(i-1,s_j)}(w_{i,j})}{\prod\limits_{j=0}^{J} \prod\limits_{i=1}^{I_j} f_{\lambda_{\widehat{\theta}}(i-1,s_j)}(w_{i,j})} \right) \leq \chi^2_{p,1-\alpha} \right\}. \tag{8}
$$

This is sometimes easier to calculate than the information-based confidence set given by (6) since no information matrix is needed. However, since the confidence set (6) is an ellipsoid, the computation of the set (8) usually takes more time than the computation of the set (6).

Remark 2. Note that the confidence sets (6) and (8) shrink to the true θ for $N := N(J) := \sum\limits_{j=0}^{J} I_j \to \infty$ if α is constant.

This happens as well if α is replaced by $\alpha(N)$ depending on N as long as $\chi^2_{p,1-\alpha(N)} \frac{1}{N}$ converges to zero. In particular, $\alpha(N)$ itself can converge to zero.

If $g(\theta)$ with $g : \Theta \to \mathbb{R}$ is a one-dimensional aspect of θ, then a classical $(1 - \alpha)$-confidence interval for $g(\theta)$ is given by

$$
\left[g(\widehat{\theta}) - q^N_{1-\alpha/2} \sqrt{\dot{\mathbf{g}}(\widehat{\theta})^\top \mathbf{I}(\widehat{\theta})^{-1} \dot{\mathbf{g}}(\widehat{\theta})}, \; g(\widehat{\theta}) + q^N_{1-\alpha/2} \sqrt{\dot{\mathbf{g}}(\widehat{\theta})^\top \mathbf{I}(\widehat{\theta})^{-1} \dot{\mathbf{g}}(\widehat{\theta})} \right], \tag{9}
$$

using the δ-method for asymptotic distributions where $\dot{\mathbf{g}}(\theta) := \frac{\partial}{\partial \theta} g(\theta)$ and q^N_α is the α-quantile of the standard normal distribution. Similarly, one-sided confidence intervals for $g(\theta)$ can be constructed by using the $(1 - \alpha)$-quantile instead of $q^N_{1-\alpha/2}$ and by replacing one of the interval boundaries with infinity.

Remark 3. Again, as in Remark 2, α may depend on the sample size $N = N(J)$ via $\alpha(N)$. Then, the confidence interval for $g(\theta)$ shrinks to the true value $g(\theta)$ for $N \to \infty$ as long as $q^N_{1-\alpha(N)/2} \sqrt{\frac{1}{N}}$ converges to zero.

For the derivation of the confidence sets (6), (8), and (9), see Appendix A.1. In particular, the confidence sets keep the nominal level of $1 - \alpha$ only asymptotically for growing sample size $N = \sum\limits_{j=0}^{J} I_j$ if some simple asymptotic requirements on the design $\mathbf{d}_J := ((1, s_0), \dots, (I_0, s_0), \dots, (1, s_J), \dots, (I_J, s_J))$ with growing sample size N hold. Note that, ideally in applications, the stress levels $s_0, \dots, s_J$ and observed numbers of failures $I_0, \dots, I_J$ should be chosen in such a way that the information matrix $\mathbf{I}(\widehat{\theta})$ leads to the smallest possible confidence set in 6. In particular, the information matrix should not be singular in order to avoid an unbounded set. For the Basquin model (4) considered here, this is ensured if there are at least observations (failures) at two different initial stress levels $s_i \neq s_j$ and preferably the distance between these two stress levels should be large. However, the maximal and minimal feasible choice for the stress levels usually depends on the underlying experiment since very small choices for s_j result in extremely long and therefore cost-inefficient failure times whereas very large choices for s_j may result in immediate failures or at least failure times that are too short to be recorded. Hence, in applications, typically an adaptive choice of the initial stress levels starting with some larger stress levels is used. In case of the application considered in this paper, the chosen initial stress levels are listed in Figure 1 in the chronological order they were used in the experiments. Note that these choices were made adaptively based on the time span of the previous experiments. In particular, we did not attempt to optimize them in terms of the resulting information matrix.

3.2 | New confidence sets for the parameter based on sign depth

Data depth is a possibility to generalize the outlier-robust median of univariate data to more complex situations. In particular, simplicial regression depth can be used for getting outlier-robust tests for hypotheses in general regression models.[38] Simplicial regression depth leads to $(p + 1)$-sign depth for many regression models if the unknown regression parameter is p-dimensional.[26] The $(p + 1)$-sign depth is the relative number of subsets with $p + 1$ observations with alternative signs

FIGURE 1 Logarithmic waiting times between the breaks of all experiments with pointwise 90%-prediction interval for the next wire break using data depth and the δ-method [Colour figure can be viewed at wileyonlinelibrary.com]
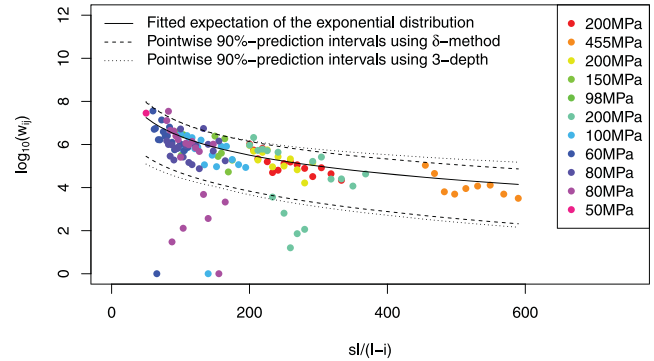
of residuals. To define sign depth here, set $N := N(J)$ and

$$\boldsymbol{w}_* := \left(w_{1,0}, \ldots, w_{I_0,0}, \ldots, w_{1,J}, \ldots, w_{I_J,J}\right)^\top \in \mathbb{R}^N,$$

$$\boldsymbol{s}_* := \left(s_0, s_0 \frac{I}{I-1}, \ldots, s_0 \frac{I}{I-(I_0-1)}, \ldots, s_J, s_J \frac{I}{I-1}, \ldots, s_J \frac{I}{I-(I_J-1)}\right)^\top \in \mathbb{R}^N.$$

To determine alternating signs of residuals, an ordering of the explanatory variables is needed. Therefore, let $\tilde{\boldsymbol{s}} := (\tilde{s}_1, \ldots, \tilde{s}_N)^\top$ be the vector obtained by sorting $\boldsymbol{s}_*$ in ascending order and let $\tilde{\boldsymbol{w}} := (\tilde{w}_1, \ldots, \tilde{w}_N)^\top$ be the corresponding vector obtained from $\boldsymbol{w}_*$ when applying the same permutation as used for $\tilde{\boldsymbol{s}}$. If some initial stress levels are equal, then some of the components of $\tilde{\boldsymbol{s}}$ are equal so that it is unclear how to order the corresponding observations in $\tilde{\boldsymbol{w}}$. Each time when this happens, competing observations are ordered randomly, independently from the other cases where this happens. Since $W_{i,j}$ are exponentially distributed with parameter $\lambda_\theta(i-1, s_j)$ for $j = 0, \ldots, J, i = 1, \ldots, I_j$, the median is $\frac{\ln(2)}{\lambda_\theta(i-1,s_j)}$. Then, the residuals

$$r_n(\boldsymbol{\theta}) := \tilde{w}_n - \frac{\ln(2)}{h_\theta(\tilde{s}_n)}, \quad n = 1, \ldots, N$$

are realizations of independent random variables $R_n(\boldsymbol{\theta})$, which have a median zero if $\boldsymbol{\theta}$ is the true parameter. Setting $\boldsymbol{z}_* = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N)^\top$ with $\boldsymbol{z}_n = (\tilde{s}_n, \tilde{w}_n)^\top$, the $(p+1)$-sign depth is defined as

$$d_S(\boldsymbol{\theta}, \boldsymbol{z}_*) = \frac{1}{\binom{N}{p+1}} \sum_{1 \le n_1 < n_2 < \ldots < n_{p+1} \le N} \left(\prod_{k=1}^{p+1} \mathbb{1}\left\{r_{n_k}(\boldsymbol{\theta})(-1)^k > 0\right\} + \prod_{k=1}^{p+1} \mathbb{1}\left\{r_{n_k}(\boldsymbol{\theta})(-1)^{k+1} > 0\right\} \right).$$

A parameter $\boldsymbol{\theta}$ is appropriate for a data set $\boldsymbol{z}_*$ if its sign depth is not smaller than the α-quantile of the distribution of the sign depth. The distribution can be determined exactly for small sample sizes since only 2^N different realizations of the signs are possible almost surely for a continuous distribution of the residuals. However, for larger samples sizes, an asymptotic distribution is needed. These asymptotic distributions are known[25,39] for $p = 1$ and for $p = 2$. They are given in these papers for autoregressive models, but they also hold for other regression models since only the signs of the residuals are used. There is ongoing research to derive the asymptotic distribution for $p > 2$. Alternatively, simplified versions of the sign depth as proposed by Kustosz et al[26] can be used for $p > 2$. However, we do not recommend using the simplified version since it usually leads to much larger confidence sets.

Having an α-quantile q_α of the distribution of the sign depth, a $(1 - \alpha)$-confidence set based on sign depth is given by

$$\{\boldsymbol{\theta}; d_S(\boldsymbol{\theta}, \boldsymbol{z}_*) \ge q_\alpha\}. \tag{10}$$

Note that tests based on $(p+1)$-sign depth are generalizations of the classical sign test since they are equivalent to the classical sign test for $p = 1$. Hence, they are outlier-robust. However, previous results[25,26] indicate that tests based on $(p+1)$-sign depth with $p \ge 2$ are much more powerful than the classical sign test.

4 | PREDICTION INTERVALS

This section contains several different approaches to construct asymptotic $(1 - \alpha)$-prediction intervals $\hat{I}_{pred}(W_*)$ for the time $W := W_{I_0+1,0} + \ldots + W_{I_c,0}$ between the I_0th and the I_cth failure based on the waiting times $W_* := (W_{1,0}, \ldots, W_{I_0,0}, \ldots, W_{1,J}, \ldots, W_{I_J,J})$ observed so far. More formally, $\hat{I}_{pred}(W_*)$ is an interval computed based on W_* such that

$$\lim_{J\to\infty} P_\theta(W \in \hat{I}_{pred}(W_*)) \geq 1 - \alpha, \quad \text{for all} \quad \theta \in \Theta. \tag{11}$$

Such prediction intervals usually require knowledge about the predictive distribution, that is, the distribution of the future random variable W. Hence, the cumulative distribution function and the quantiles of W are given in Section 4.1. Afterwards, the following prediction intervals are presented:

(a) Naive/Plug-in prediction intervals (14) based on a maximum likelihood estimator for θ (Section 4.2).
(b) Prediction intervals (15) based on a confidence set for θ computed in one of the following ways (Section 4.3):

- Using the Wald test (6).
- Using the likelihood-ratio test (8).
- Using the outlier-robust sign depth (10).

(c) Prediction intervals (19) based on confidence intervals (9) for the quantiles of the predictive distribution (Section 4.4).

Each prediction interval has its own advantages and disadvantages, which are highlighted in Sections 5 and 6. Most notably, approach (a) provides the smallest prediction intervals but struggles to keep the nominal $(1 - \alpha)$-coverage rate unless the samples size is very large and no outliers occur in the data. Approach (b) based on the sign depth is the only outlier-robust prediction presented in this paper and thus outperforms all other approaches in terms of coverage rate whenever outliers appear frequently in the data. However, this approach often produces the largest prediction intervals and can lead to very conservative predictions, in particular in applications where outlier-robustness is not needed. Finally, approach (c) is computationally way more efficient than (b) and also yields smaller prediction intervals while still keeping the nominal $(1 - \alpha)$-coverage rate in applications without outliers.

4.1 | Predictive distribution of the failure times

If all $W_{I_0+1,0}, \ldots, W_{I_c,0}$ had the same exponential distribution, then W would have a Gamma distribution and prediction intervals could be constructed via known results for Gamma distributions. However, here $\lambda_\theta(I_0, s_0), \ldots, \lambda_\theta(I_c - 1, s_0)$ are pairwise different, and therefore W has a hypoexponential distribution with density given by[46]

$$f_{W,\theta}(w) := \sum_{i=I_0}^{I_c-1} \lambda_\theta(i, s_0) \, e^{-w \, \lambda_\theta(i,s_0)} \, a_i(\theta) \quad \text{with} \quad a_i(\theta) := \prod_{k=I_0, k\neq i}^{I_c-1} \frac{\lambda_\theta(k, s_0)}{\lambda_\theta(k, s_0) - \lambda_\theta(i, s_0)}.$$

Hence, the cumulative distribution function is

$$F_{W,\theta}(w) := \sum_{i=I_0}^{I_c-1} a_i(\theta) \left(1 - e^{-w \, \lambda_\theta(i,s_0)}\right). \tag{12}$$

Expression (12) can be computed efficiently, see R package `sdprisk`.[47] However, this implementation suffers from numerical instability when $\lambda_\theta(I_0, s_0), \ldots, \lambda_\theta(I_c - 1, s_0)$ do not differ much.[48] Since this is often the case in our data, we need a numerically more stable algorithm. This can be done by rewriting the cumulative distribution function using a matrix exponential as follows:

$$F_{W,\theta}(w) = 1 - e_1^\top \exp(Dw)\mathbf{1}_{I_c-I_0} = 1 - e_1^\top \sum_{k=0}^{\infty} \frac{D^k w^k}{k!} \mathbf{1}_{I_c-I_0}, \tag{13}$$

where $e_1 = (1, 0, \ldots, 0)^\top$ is the first $(I_c - I_0) \times 1$ unit vector, $\mathbf{1}_{I_c - I_0} = (1, 1, \ldots, 1)^\top$ is an $(I_c - I_0) \times 1$ vector of ones, and

$$D = \begin{pmatrix} -\lambda_\theta(I_0, s_0) & \lambda_\theta(I_0, s_0) & 0 & \ldots & 0 \\ 0 & -\lambda_\theta(I_0 + 1, s_0) & \lambda_\theta(I_0 + 1, s_0) & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & -\lambda_\theta(I_c - 2, s_0) & \lambda_\theta(I_c - 2, s_0) \\ 0 & \ldots & 0 & 0 & -\lambda_\theta(I_c - 1, s_0) \end{pmatrix}$$

is a $(I_c - I_0) \times (I_c - I_0)$ matrix. An implementation[23] of (13) is considerably more stable than the direct computation of expression (12). The infinite series $\sum_{k=0}^\infty \frac{D^k w^k}{k!} \mathbf{1}_{I_c - I_0}$ is computed via the `Matrix` R package[49] using Ward's diagonal Padé approximation.

An α-quantile $b_\alpha(\theta)$ of this distribution can only be given implicitly, namely as

$$H_\alpha(\theta, b_\alpha(\theta)) = 0 \quad \text{where} \quad H_\alpha(\theta, b) := \sum_{i=I_0}^{I_c - 1} a_i(\theta) \left(1 - e^{-b \, \lambda_\theta(i, s_0)}\right) - \alpha$$

or equivalently

$$H_\alpha(\theta, b) = 1 - e_1^\top \sum_{k=0}^\infty \frac{D^k b^k}{k!} \mathbf{1}_{I_c - I_0} - \alpha.$$

4.2 | Background and naive approach

Although there are many methods for getting prediction intervals for exponential and other univariate lifetime distributions,[2,40-43] not much has been published for accelerated lifetime experiments. Patel[2] mentions only a prediction interval for accelerated lifetime experiments with normal distribution, while Xiong and Milliken[44] provide simulation-based prediction intervals for exponentially distributed lifetimes for accelerated life testing. Hong and Meeker[45] provide a prediction method based on dynamic covariate information, but no prediction intervals are given.

Our (simulation-free) prediction intervals all require knowledge about the predictive distribution analyzed in Section 4.1, that is, the distribution of the random variable W in (11). If the α-quantiles of this predictive distribution are denoted by $b_\alpha(\theta)$, then the so-called naive or plug-in prediction interval for W is given by

$$\left[b_{\alpha/2}(\widehat{\theta}), b_{1-\alpha/2}(\widehat{\theta})\right], \tag{14}$$

where $\widehat{\theta} = \widehat{\theta}(W_*)$ is any consistent estimator for θ such as the maximum likelihood estimator. This prediction interval is popular among engineers due to its simplicity. While it yields good results for large sample sizes, small sample sizes can lead to prediction intervals with a poor coverage rate.[50] Therefore, we propose four more conservative approaches for prediction intervals in Sections 4.3 and 4.4 below. These approaches can also be used in other situations where the predicted variable W is independent of the observed random vector W_*.

4.3 | Prediction intervals based on confidence sets for the model parameter

The first approach is based on confidence sets for θ like those given by (6), (8), and (10). Throughout the section, $\alpha(1)$ and $\alpha(2)$ denote given error rates in $(0, 1)$, representing the choice for α in the confidence set and the choice for α in the quantiles of the predictive distribution. Note that we do not use the more common subscript notation α_1 and α_2 to avoid double or triple subscripts in the formulas.

Theorem 1. *If $\hat{C}_{\alpha(1)}(W_*)$ is an asymptotic $(1 - \alpha(1))$-confidence set for θ, then the interval $\hat{I}_{pred}^1(W_*)$ given by*

$$\hat{I}_{pred}^1(W_*) := \bigcup_{\theta \, \in \hat{C}_{\alpha(1)}(W_*)} \left[b_{\alpha(2)/2}(\theta), b_{1-\alpha(2)/2}(\theta)\right] \tag{15}$$

is an asymptotic $(1 - \alpha(1))(1 - \alpha(2))$-prediction interval for $W = W_{I_0+1,0} + \ldots + W_{I_c,0}$.

For the choice of $\alpha(1)$, providing the confidence set, and $\alpha(2)$, providing the interval for the future observation, see Remark 4. Note that it is sometimes easier to work with the following larger prediction interval:

$$\hat{I}^1_{pred}(W_*) \subseteq \left[\min\{b_{\alpha(2)/2}(\boldsymbol{\theta}); \boldsymbol{\theta} \in \hat{C}_{\alpha(1)}(W_*)\}, \max\{b_{1-\alpha(2)/2}(\boldsymbol{\theta}); \boldsymbol{\theta} \in \hat{C}_{\alpha(1)}(W_*)\}\right]. \tag{16}$$

All prediction intervals (based on confidence sets) in this article are computed via (16) rather than (15) to reduce the computational effort. Note that both prediction intervals often coincide, for example, when the confidence set $\hat{C}_{\alpha(1)}(W_*)$ is connected and the mapping $\boldsymbol{\theta} \mapsto b_\alpha(\boldsymbol{\theta})$ is continuous for every α.

4.4 | Prediction intervals based on confidence intervals for quantiles

The construction of a confidence set $\hat{C}_{\alpha(1)}(W_*)$ for $\boldsymbol{\theta}$ is sometimes not that easy, in particular for high-dimensional parameters $\boldsymbol{\theta}$. Therefore, it might be easier to base the prediction interval on confidence intervals of the quantiles $b_\alpha(\boldsymbol{\theta})$ of the predictive distribution. Since these confidence intervals are obtained with the δ-method for asymptotic distributions, we call such prediction intervals prediction intervals by the δ-method. To this end, let $\dot{\boldsymbol{b}}_\alpha(\boldsymbol{\theta}) := \frac{\partial}{\partial\theta} b_\alpha(\boldsymbol{\theta})$. The implicit function theorem yields

$$\dot{\boldsymbol{b}}_\alpha(\boldsymbol{\theta}) = -\left(\frac{\partial}{\partial\tilde{b}} H_\alpha(\tilde{\theta},\tilde{b})\Big|_{(\tilde{\theta},\tilde{b})=(\theta,b_\alpha(\theta))}\right)^{-1} \frac{\partial}{\partial\tilde{\theta}} H_\alpha(\tilde{\theta},\tilde{b})\Big|_{(\tilde{\theta},\tilde{b})=(\theta,b_\alpha(\theta))},$$

which can be calculated explicitly. Setting $g(\boldsymbol{\theta}) = b_{\alpha(2)/2}(\boldsymbol{\theta})$ and $g(\boldsymbol{\theta}) = b_{1-\alpha(2)/2}(\boldsymbol{\theta})$, respectively, Section 3.1 ensures that

$$\left[b_{\alpha(2)/2}(\widehat{\boldsymbol{\theta}}) - \hat{v}_1, \infty\right), \quad \hat{v}_1 := q^N_{1-\alpha(1)/2} \sqrt{\dot{\boldsymbol{b}}_{\alpha(2)/2}(\widehat{\boldsymbol{\theta}})^\top \mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1} \dot{\boldsymbol{b}}_{\alpha(2)/2}(\widehat{\boldsymbol{\theta}})} \tag{17}$$

and

$$\left(-\infty, b_{1-\alpha(2)/2}(\widehat{\boldsymbol{\theta}}) + \hat{v}_2\right], \quad \hat{v}_2 := q^N_{1-\alpha(1)/2} \sqrt{\dot{\boldsymbol{b}}_{1-\alpha(2)/2}(\widehat{\boldsymbol{\theta}})^\top \mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1} \dot{\boldsymbol{b}}_{1-\alpha(2)/2}(\widehat{\boldsymbol{\theta}})}, \tag{18}$$

are one-sided $\left(1 - \frac{\alpha(1)}{2}\right)$-confidence intervals for $b_{\alpha(2)/2}(\boldsymbol{\theta})$ and $b_{1-\alpha(2)/2}(\boldsymbol{\theta})$, respectively.

Theorem 2. *Let $\hat{v}_1$ and $\hat{v}_2$ be as in (17) and (18). Then, the interval $\hat{I}^2_{pred}(W_*)$ given by*

$$\hat{I}^2_{pred}(W_*) := \left[b_{\alpha(2)/2}(\widehat{\boldsymbol{\theta}}) - \hat{v}_1, b_{1-\alpha(2)/2}(\widehat{\boldsymbol{\theta}}) + \hat{v}_2\right] \tag{19}$$

is an asymptotic $(1 - \alpha(1))(1 - \alpha(2))$-prediction interval for $W = W_{I_0+1,0} + \ldots + W_{I_c,0}$.

Note that $b_\alpha(\boldsymbol{\theta})$ can be given explicitly if $I_c = I_0 + 1$, that is, if we want to predict only the next component failure. Then, we have

$$b_\alpha(\boldsymbol{\theta}) = -\frac{\ln(1-\alpha)}{\lambda_\theta(I_0, s_0)} = \frac{\ln\left(\frac{1}{1-\alpha}\right)}{\lambda_\theta(I_0, s_0)} = \ln\left(\frac{1}{1-\alpha}\right) g(\boldsymbol{\theta})$$

with $g(\boldsymbol{\theta}) = (\lambda_\theta(I_0, s_0))^{-1}$. By choosing

$$\hat{v} := q^N_{1-\alpha(1)/2} \sqrt{\dot{\mathbf{g}}(\widehat{\boldsymbol{\theta}})^\top \mathbf{I}(\widehat{\boldsymbol{\theta}})^{-1} \dot{\mathbf{g}}(\widehat{\boldsymbol{\theta}})},$$

one obtains a $(1 - \alpha(1))(1 - \alpha(2))$-prediction interval $\hat{I}^2_{pred}(W_*)$ via

$$\hat{I}^2_{pred}(W_*) = \left[\ln\left(\frac{2}{2-\alpha(2)}\right) \left(g(\widehat{\boldsymbol{\theta}}) - \hat{v}\right), \ln\left(\frac{2}{\alpha(2)}\right) \left(g(\widehat{\boldsymbol{\theta}}) + \hat{v}\right)\right].$$

Remark 4. All approaches of prediction intervals in Sections 4.3 and 4.4 lead to $(1-\alpha(1))(1-\alpha(2))$-prediction intervals. If the level of the prediction interval should be $1-\alpha$, then $1-\alpha = (1-\alpha(1))(1-\alpha(2))$ must be satisfied. For small sample sizes, a good choice is $\alpha(1) = \alpha(2)$ so that $1-\alpha = (1-\alpha(1))^2$. For larger sample sizes, it is better to choose $\alpha(1)$ smaller than $\alpha(2)$ since confidence sets shrink to the true parameter if the sample size goes to infinity. As noted in the remarks in Section 3.1, this may happen as well if $\alpha(1)$ converges to zero for growing sample size. Since $1-\alpha = (1-\alpha(1))(1-\alpha(2))$ implies $\alpha(2) = 1 - \frac{1-\alpha}{1-\alpha(1)}$, a convergence of $\alpha(1)$ to zero means that $\alpha(2)$ converges to α. In such situations, the proposed prediction intervals behave asymptotically like the so called naive or plug-in prediction interval given by (14).

Remark 5. All confidence set based approaches typically lead to conservative prediction intervals, that is, the true coverage rate is usually above the $(1 - \alpha)$ level the predictions are designed for. This is not surprising since these prediction intervals will have at least a $(1-\alpha(2))$-coverage rate[†] whenever the true parameter lies within the confidence set. Even if this is not the case, the coverage rate will often be fairly large simply due to the large intervals provided. The same is also true for predictions based on the δ-method although these prediction intervals are usually a bit smaller.

The naive approach (14) is recommended if the sample size is sufficiently large[‡] and/or nonconservative predictions are not an issue. However, aside from having small sample sizes, we suggest to use the confidence-set-based predictions/δ-method in the following situations:

- The prediction model is an oversimplification of the true dynamics in the experiment. This is often the case and can cause the naive approach to yield poor coverage rates. Our application in the next section is indeed an example for a poor coverage rate.
- The data can be contaminated with outliers. If some experiments might yield unexpectedly short/long component lifetimes, due to, for example, fabrication errors in the components, we recommend using a confidence set based on an outlier-robust method, for example, the approach in Section 3.2. A simulation study with outlier contaminated data can be found in Section 6.

Note that the approach in Section 3.2 can lead to an extremely large confidence set (and thus also prediction intervals) if the sample size is small. We therefore recommend this approach only if a sufficiently large number of waiting times (> 100) is observed.

5 | APPLICATIONS

In this section, we analyze data from J=11 experiments with concrete beams. Each beam contains a total of $I = 35$ tension wires and is exposed to cyclic loading. During this process, the tension wires successively break. Since wire breaks produce a loud noise, the failure times can be determined exactly using acoustic measurements. Throughout the analysis, we consider the Basquin model given by the link function $h_\theta(x) = \exp(-\theta_1 + \theta_2 \cdot \ln(x))$ since initial investigations indicated that this is the best choice among the link functions we considered. The initial stress levels $s_1, \ldots, s_{11}$ used in the experiments and a plot of the observed waiting times w_{ij} between the breaks against the actual loads of $\frac{s_j I}{I-i}$ are presented in Figure 1. Before applying our prediction methods, we start with the model check given in Section 2.3.

5.1 | Model check

The confidence band of the Q-Q plot in Figure 2 is computed by the simulated simultaneous 90%-confidence band presented in Section 2.3 using $M = 1\,000\,000$, that is, one million samples, from our model with a parameter θ equal to the MLE $\hat{\theta}$ of the data example. Although it seems like all quantiles lie within the confidence band, actually the first few sample quantiles are slightly below it. This is due to some lower outliers, see Figure 1 for a plot of all waiting times on a logarithmic scale. Ideally, one should therefore choose a robust parameter estimation/confidence set, for example, the method presented in Section 3.2.

5.2 | Comparison of the prediction intervals via cross validation

We now apply the different prediction methods from Section 4 to the data from the concrete beam experiments. Recall that we consider the Basquin model, that is, the link function $h_\theta(x) = \exp(-\theta_1 + \theta_2 \cdot \ln(x))$. Also note that since the unknown parameter vector θ is two-dimensional, the sign depth from Section 3.2 is based on subsets with three residuals and therefore called 3-depth. The following predictions intervals are compared:

[†]usually even more than that since the interval is larger than $[b_{\alpha(2)/2}(\theta), b_{1-\alpha(2)/2}(\theta)]$
[‡]The simulation study in Section 6 suggests that naive predictions based on around 150 observed waiting times already provide intervals with a coverage rate close to $1 - \alpha$.
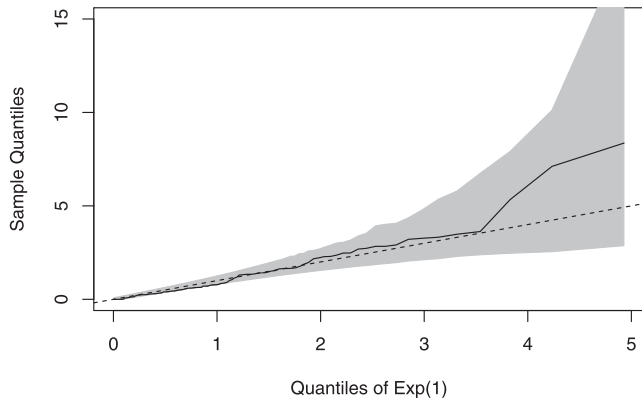
FIGURE 2 Q-Q plot of the (linearly interpolated) quantiles of the rescaled data compared with the standard exponential distribution. The grey area represents a simultaneous 90%-confidence band for the (rescaled) exponential distribution, see (3) for the rescaling

Wald: information based prediction interval given by (15) based on (6),

LR: likelihood ratio based prediction interval given by (15) based on (8),

3-depth: 3-depth prediction interval given by (15) based on (10),

δ-method: δ-method based prediction interval given by (19),

Naive: the naive/plug-in prediction interval in (14) using a MLE $\hat{\theta}$ for θ.

For all prediction intervals, a level of 90% was chosen, which is a common choice for applications in construction engineering. That means $\alpha(1) = \alpha(2) \approx 0.0513$ for the methods Wald, LR, 3-depth, and δ-method to ensure the level $(1 - \alpha(1))(1 - \alpha(2)) = 90\%$.

Figure 1 shows the logarithmic waiting times between the breaks of all experiments together with the estimated function $\log_{10}(h_{\hat{\theta}}(\cdot)^{-1})$ using the maximum likelihood estimator $\hat{\theta} = (27.99, 2.89)$ and the pointwise prediction intervals for the logarithmic waiting time of the next break based on the two new prediction methods 3-depth and δ-method. The majority of the data is well approximated by our estimations. The δ-method mostly provides smaller prediction intervals than the 3-depth method, the only exception is the upper bound for small stress ranges, in which both methods yield nearly the same bound. However, the intervals are very similar and, in particular, not much influenced by the outliers with very small waiting times.

A comparison of all prediction methods is done by a leave-one-out cross-validation. The following scenarios for leave-one-out cross-validation are considered:

Next: For all $j = 1, \ldots, J = 11$ and $L = 0, \ldots, I_j - 1$,
the waiting time $W_{L+1,j}$ for the next event is predicted using
$w_{1,j}, \ldots, w_{L,j}$ and $w_{i,l}$ for $i = 1, \ldots, I_l, \ l \neq j$,

Next 5: For all $j = 1, \ldots, J = 11$ and $L = 0, \ldots, I_j - 5$,
the waiting time $w_{1,j} + \ldots + w_{L,j} + W_{L+1,j} + \ldots + W_{L+5,j}$
until the future five events is predicted using
$w_{1,j}, \ldots, w_{L,j}$ and $w_{i,l}$ for $i = 1, \ldots, I_l, \ l \neq j$.

The chosen $j = 1, \ldots, J = 11$ takes the role of the series that is to be predicted, and L denotes the number of wire breaks already observed in this series, that is, $L = I_0$ in the setup described in (2). Again, $L = 0$ means that no observations of the new series are used.

To measure the performance of the prediction intervals, the interval score[51] is used. It is given by

$$S([l, u], w_0) := (u - l) + \frac{2}{\alpha}(l - w_0)\mathbb{1}\{w_0 < l\} + \frac{2}{\alpha}(w_0 - u)\mathbb{1}\{w_0 > u\}$$

if $[l, u]$ is the prediction interval, w_0 is the realized future observation, and $\mathbb{1}\{\ \}$ denotes the indicator function. It combines the coverage and length of the prediction interval in such a way that a small score corresponds to a good prediction interval.

TABLE 1 Median interval score in millions and coverage rates for the 90%-prediction intervals in the scenarios Next and Next 5 for the different prediction methods

	Interval score		Coverage rate	
Method	Next	Next 5	Next	Next 5
3-depth	11.24	25.78	0.89	0.98
δ-method	11.10	24.29	0.88	0.91
Wald	11.71	25.14	0.88	0.92
LR	11.71	25.26	0.88	0.92
Naive	9.62	26.99	0.83	0.74
Sample size	137	97	137	97

FIGURE 3 A, 90%-prediction intervals for a new experiment SB06 with a very low stress range of 50 MPa. The arrow indicates that the experiment was stopped after 108 million load cycles with only one wire breakage. B, The restarted experiment SB06a with a higher stress range of 120 MPa [Colour figure can be viewed at wileyonlinelibrary.com]
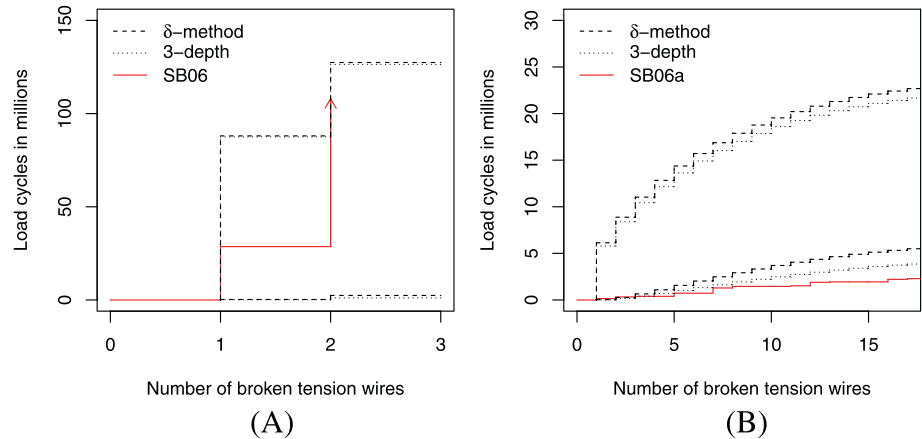
Table 1 shows the average value of this score in millions and the coverage rates in the two scenarios. Note that the naive method struggles to reach the 90%-coverage rate since it provides too small prediction intervals. This leads to the best interval score in the scenario Next and the worst interval score in scenario Next 5. The other methods provide coverage rates close to or above the nominal level of 90%. The δ-method leads to the best interval score within these methods. Since the 3-depth method provides larger prediction intervals, its interval score is worse. However, it still beats the Wald and LR methods in scenario Next. In scenario Next 5, its prediction intervals are so large that the coverage is much above 90%, causing its interval score to be the worst among the non-naive methods.

5.3 | The prediction intervals in the accelerated set-up

Finally, we discuss predictions for the last experiment with a lower stress range and only one wire break in total. Figure 3A shows the 90%-prediction intervals for this experiment called SB06 with a stress range $s_0 = 50$ MPa using the depth method and the δ-method based on the $J = 10$ experiments with stress ranges from 80 to 455 MPa. Again, both methods provide very similar prediction intervals where lower and upper bounds of the prediction intervals of the 3-depth method are a little bit smaller than those of the δ-method. Both prediction intervals cover the first wire breakage at about 28 million load cycles given by the red line. The experiment was stopped after about 108 million load cycles without observing a second wire failure. This results in a censored observation indicated by the arrow in Figure 3A. Both methods predict another wire break within the next 25 million load cycles, hence indicating that the experiment should have been run a little longer to produce more data.

Remark 6. In order to check for preexisting defects in the tension wires caused by the 108 million load cycles at 50 MPa, the experiment was restarted at a higher stress range of 120 MPa as SB06a. The course of this experiment is depicted in Figure 3B along with the 90%-prediction intervals given by the two methods where no preexisting damage is assumed, that is, the predictions in Figure 3B are done without using the results for SB06 at 50 MPa (note that this extra failure time would have barely any effect on the overall prediction). The prediction intervals are again similar but do not include the true failure times of the experiment SB06a. The wires broke earlier than our methods proposed. This makes preexisting defects in the wires plausible.
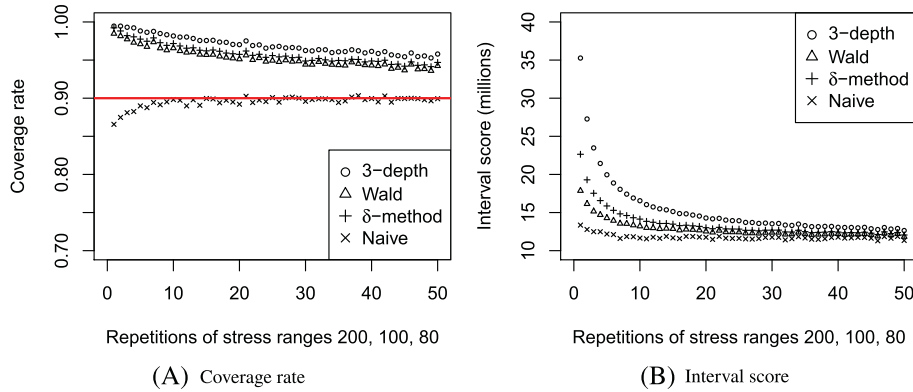
FIGURE 4 A, Coverage rates and B, scores of different methods applied to simulated data without contamination [Colour figure can be viewed at wileyonlinelibrary.com]

6 | SIMULATION STUDY

We consider a setup where several wire breaks in models with high stress levels are observed (representing the accelerated lifetime experiments), which then are used to make predictions for a system with a fairly low stress level. More precisely, we simulate the first six wire breaks (i.e., $I_j = 6$ for all j) in experiments with a total number of $I = 10$ wires[§] based on the Basquin link function with $\theta = (28, 3)$ (which corresponds to the rounded MLE in the real data). The initial stress ranges for the observed experiments are a K-times repetition of the triple $s = 200, 100, 80$ for several integers K. The prediction interval concerns the sixth break in an experiment with initial stress range $s_0 = 60$ where only the first three wire breaks are taken as part of the observed data. The overall prediction error is set to be $\alpha = 0.1$. As suggested in the remark at the end of Section 4, this error is split into

$$\alpha_1(K) = \frac{0.05}{\sqrt{K}}, \qquad \alpha_2(K) = 1 - \frac{1 - 0.1}{1 - \alpha_1(K)},$$

in the cases of the δ-method and methods based on confidence sets, where $(\alpha_1(K), \alpha_2(K))$ represent the errors $(\alpha(1), \alpha(2))$ occurring in Sections 4.3 and 4.4. Note that the choice for $\alpha_1(K)$ is somewhat arbitrary among choices that ensure $\alpha_2(K) \to \alpha$ as $K \to \infty$ and that also lead to confidence sets that converge to the one-point set $\{\theta\}$ as $K \to \infty$.

The simulation results for $K = 1, \dots, 50$ (averaged over 10,000 simulations) are given in Figure 4. Although the naive approach struggles to keep the 90%-coverage rate if the dataset is fairly small, it still yields better results than the other approaches in terms of interval score (note that a small score corresponds to a good prediction interval). This is not surprising given that the other prediction methods are more conservative and lead to larger prediction intervals with an unnecessarily high coverage rate.

Contaminated data. To study the behavior of the methods in the presence of outliers, we contaminated the data as follows: Each of the $18K$ simulated waiting times for $s = 200, 100, 80$ has a 20% chance to become an outlier (the choices are made independently from each other). Each chosen waiting time is replaced by an outlier where with the probability 50% the outlier is given by the 10th of the original value and with probability 50% it is set to 1. Outliers with small values appear often in applications as can be seen from the data set in Figure 1. However, it is known that methods for the exponential distribution are fairly robust against small outliers since these outliers are bounded by zero. This is also visible from the results in Section 5.2. But there is also the phenomenon of fatigue-tested specimen without rupture at stress levels where this is not expected. This means that sometimes unusual long waiting times appear. These are the outliers with large values.

Figure 5 shows the results for the contaminated data. Now the naive method always has the problem that its coverage rate is much smaller than the nominal level. This also happens for the other likelihood-based methods Wald, LR, and δ-method although the deviation from the nominal coverage level is not as large as for the naive method. Only the 3-depth method keeps the level so that its score becomes the best for larger sample sizes.

Similar results are obtained with other scenarios of contamination, see Supporting Information.

[§]we reduced the total number of wires in order to obtain a relative number $I_j/I = 0.6$ of broken wires which is similar to the number of wire break in, e.g, experiment SB04 in the real data

FIGURE 5 A, Coverage rates and B, scores of the different methods applied to contaminated data. The contamination chance is 20%. Contaminated data has a 50% chance to increase the waiting time by a factor of 10 and a 50% chance to decrease the waiting time to 1 [Colour figure can be viewed at wileyonlinelibrary.com]
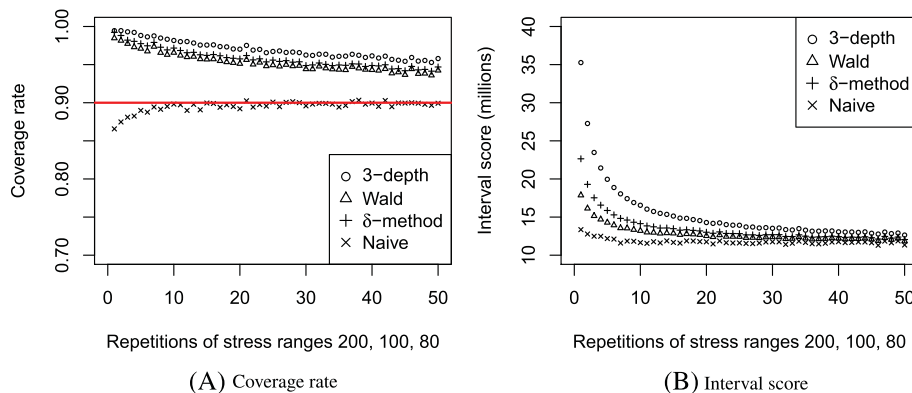
7 | CONCLUSION

We have proposed two new prediction intervals for the time of the failure of the I_cth component of a load-sharing system with I components when the failure process of other systems with I components under different stress levels has been observed before. One of the new prediction intervals is outlier-robust because it is only based on the signs of residuals of the observed failures. Outlier-robustness is important since sometimes the waiting times between the failure of components are very short. Also, unusually long waiting times appear in so-called fatigue tests without rupture. Although outlier-robust methods are often less efficient than classical methods, the application to a real data set coming from experiments with prestressed concrete beams showed that they behave similarly or even better than the classical methods based on the Wald test and the likelihood ratio test. In the simulation study, they become superior for data with contamination.

The outlier-robust methods and the classical methods have the disadvantage that confidence sets for the unknown parameter must be calculated. In particular, this can be very time consuming for higher-dimensional parameters and usually has to be approximated by, for example, a grid search. Therefore, we proposed a more accurate method via the δ-method and the implicit function theorem where no confidence sets of the parameters have to be calculated. The application to the real data set and the simulation study show that this method yields fairly good results although the used dimension of the parameter was not high. It can be assumed that the superiority of this method compared with the confidence-set-based approaches becomes higher if the parameter has a higher dimension.

We assumed that the number I of components is the same for all systems since this was the case in our application with experiments with prestressed concrete beams. An extension to systems with different numbers of components can be done easily. We ignored censoring since most of our experiments were stopped after the last reported failure. Censoring can be easily incorporated in the likelihood based methods, but for the method based on sign depth, it is still unclear how this can be done. Similarly, an extension to a model incorporating damage accumulation and aging of the components will be much more complicated. Then, simple state-dependent processes are not adequate anymore and processes from the more general class of self-exciting processes must be used.

Self-exciting point processes include many point processes often used in reliability analysis as the homogeneous or inhomogeneous Poisson process or renewal processes.[27,30,52] Although a combination of a nonhomogeneous Poisson process with a birth process would be very adequate to model damage accumulation and load redistribution, we are not aware of a statistical analysis for such models. The only related paper[53] considers a combination of (nonhomogeneous) Poisson processes and renewal processes. In this paper, Lawless and Thiagarajah derive maximum likelihood estimators and apply a Wald test for a nonhomogeneous Poisson process of airplane air-condition failure data. Dachian and Kutoyants[54] also developed a test for testing a stationary Poisson process against a stationary self-exciting point processes. Other extensions of the nonhomegeneous Poisson process and renewal processes can be found in the literature.[55,56]

Statistical inference for the more general class of self-exciting processes is treated mainly for other fields of application as market analysis,[57] analysis of financial data,[58-62] neuron firing data,[63] cancer data,[64] or earthquake data.[65] These papers treat estimation and testing of unknown parameters of the model but not prediction intervals. Only one[58] of the papers provides prediction intervals, namely for the terminal prices in online auctions. However, the prediction intervals are obtained by simulations.

We expect that simulation-free predictions intervals will be difficult to derive for self-exciting point processes including load sharing and damage accumulation. This is mainly due to the fact that our simulation-free prediction intervals depend heavily on knowing the distribution of the sum of the waiting times, that is, the hypoexponential distribution in our

approach, explicitly. Without assuming independence between waiting times, it is usually very hard to compute this distribution analytically and simulation-based predictions are required.

ORCID

Kevin Leckey https://orcid.org/0000-0002-4714-9430

Christine H. Müller https://orcid.org/0000-0002-4097-3320

REFERENCES

1. Kvam PH, Peña EA. Estimating load-sharing properties in a dynamic reliability system. *J Am Stat Assoc*. 2005;100(469):262-272. https://doi.org/10.1198/016214504000000863

2. Patel JK. Prediction intervals—a review. *Comm Stati Theory and Methods*. 1989;18(7):2393-2465. https://doi.org/10.1080/03610928908830043

3. Hsieh H. Prediction intervals for Weibull observations, based on early-failure data. *IEEE Trans Reliab*. 1996;45(4):666-670. https://doi.org/10.1109/24.556591

4. Barakat H, El-Adll ME, Aly AE. Exact prediction intervals for future exponential lifetime based on random generalized order statistics. *Comput Math Appl*. 2011;61(5):1366-1378. https://doi.org/10.1016/j.camwa.2011.01.002

5. Nordman DJ, Meeker WQ. Weibull prediction intervals for a future number of failures. *Technometrics*. 2002;44(1):15-23. https://doi.org/10.1198/004017002753398191

6. Cramer E, Kamps U. Sequential order statistics and *k*-out-of-*n* systems with sequentially adjusted failure rates. *Ann Inst Stat Math*. 1996;48(3):535-549. https://doi.org/10.1007/BF00050853

7. Burkschat M. Systems with failure-dependent lifetimes of components. *J Appl Probab*. 2009;46(4):1052-1072. https://doi.org/10.1239/jap/1261670688

8. Beutner E. Nonparametric model checking for *k*-out-of-*n* systems. *J Stat Plann Inference*. 2010;140(3):626-639. https://doi.org/10.1016/j.jspi.2009.08.005

9. Balakrishnan N, Beutner E, Kamps U. Modeling parameters of a load-sharing system through link functions in sequential order statistics models and associated inference. *IEEE Trans Reliab*. 2011;60(3):605-611.

10. Sutar SS, Naik-Nimbalkar UV. Accelerated failure time models for load sharing systems. *IEEE Trans Reliab*. 2014;63(3):706-714.

11. Bedbur S, Burkschat M, Kamps U. Inference in a model of successive failures with shape-adjusted hazard rates. *Ann Inst Stat Math*. 2016;68(3):639-657.

12. Kim H, Kvam PH. Reliability estimation based on system data with an unknown load share rule. *Lifetime Data Analy*. 2004;10(1):83-94. https://doi.org/10.1023/B:LIDA.0000019257.74138.b6

13. Phoenix SL, Newman WI. Time-dependent fiber bundles with local load sharing. II. General Weibull fibers. *Phys Rev E*. 2009;80:066115.

14. Park C. Parameter estimation for the reliability of load-sharing systems. *IIE Trans*. 2010;42(10):753-765. https://doi.org/10.1080/07408171003670991

15. Kong Y, Ye Z. Interval estimation for k-out-of-n load-sharing systems. *IISE Trans*. 2017;49(3):344-353. https://doi.org/10.1080/0740817X.2016.1217102

16. Xu J, Hu Q, Yu D, Xie M. Reliability demonstration test for load-sharing systems with exponential and Weibull components. *PLOS ONE*. 2017;12(12):e0189863.

17. Xu J, Xie M, Hu Q. Reliability assessment for load-sharing systems with exponential components using statistical expansion as a correction. *Appl Stoch Model Bus Ind*. 2019;35(4):998-1010. https://doi.org/10.1002/asmb.2439

18. Park C. Parameter estimation from load-sharing system data using the expectation-maximization algorithm. *IIE Trans*. 2013;45(2):147-163. https://doi.org/10.1080/0740817X.2012.669878

19. Deshpande JV, Dewan I, Naik-Nimbalkar UV. A family of distributions to model load sharing systems. *J Stat Plann Inference*. 2010;140(6):1441-1451.

20. Kong Y, Ye Z. A cumulative-exposure-based algorithm for failure data from a load-sharing system. *IEEE Trans Reliab*. 2016;65(2):1001-1013.

21. Kong Y, Ye Z. A conditional test for the exponential distribution in load-sharing systems. In: 2016 IEEE International Conference on Industrial Engineering and Engineering Management. IEEM; 2016:512-515.

22. Singh B, Sharma K, Kumar A. Aclassical and Bayesian estimation of a *k*-components load-sharing parallel system. *Comput Stat Data Anal*. 2008;52(12):5175-5185. https://doi.org/10.1016/j.csda.2008.05.026

23. Gertsbakh I, Neuman E, Vaisman R. Monte Carlo for estimating exponential convolution. *Commu Stat Simul Comp*. 2015;44(10):2696-2704. https://doi.org/10.1080/03610918.2013.842591

24. Mood AM, Graybill FA. *Introduction to the Theory of Statistics*. 3rd. New York: McGraw-Hill; 1974.

25. Kustosz CP, Leucht A, Müller CH. Tests based on simplicial depth for AR(1) models with explosion. *J Time Ser Anal*. 2016;37(6):763-784. https://doi.org/10.1111/jtsa.12186

26. Kustosz CP, Müller CH, Wendler M. Simplified simplicial depth for regression and autoregressive growth processes. *J Stat Plann Inference*. 2016;173:125-146.

27. Snyder DL, Miller MI. *Random Point Processes in Time and Space*. 2nd. New York: Springer-Verlag; 1991.

28. Jacobsen M. *Point Process Theory and Applications. Marked Point and Piecewise Deterministic Processes*. Boston: Birkhäuser; 2006.

29. Basquin O. The exponential law of endurance tests. *American Soc Test Mater Proceed*. 1910;10:625-630.

30. Sobczyk K, Spencer BF. *Random Fatigue. From Data to Theory*. Boston: Academic Press; 1992.

31. Xu J, Liu B, Zhao X. Parameter estimation for load-sharing system subject to Wiener degradation process using the expectation-maximization algorithm. *Qual Reliab Eng Int*. 2019;35(4):1010-1024. https://doi.org/10.1002/qre.2442

32. Cinlar E. *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall; 1975.

33. Henze N, Meintanis SG. Recent and classical tests for exponentiality: a partial review with comparisons. *Metrika*. 2005;61(1):29-45. https://doi.org/10.1007/s001840400322

34. David FN, Johnson N. The probability integral transformation when parameters are estimated from the sample. *Biometrika*. 1948;35(1/2):182-190. https://doi.org/10.2307/2332638

35. Schervish MJ. *Theory of Statistics*. New York: Springer-Verlag; 1995.

36. Müller CH. D-optimal designs for lifetime experiments with exponential distribution and censoring. In: Ucinski D, Atkinson AC, Patan M, eds. *mODa 10 - Advances in Model-Oriented Design and Analysis. Contributions to Statistics*. Heidelberg: Springer International Publishing; 2013:179-186.

37. Rao CR. *Linear Statistical Inference and its Applications*. New York-London-Sydney: John Wiley and Sons, Inc. XVIII; 1965.

38. Müller CH. Depth estimators and tests based on the likelihood principle with application to regression. *J Multivariate Anal*. 2005;95(1):153-181. https://doi.org/10.1016/j.jmva.2004.06.006

39. Kustosz CP, Müller CH. Analysis of crack growth with robust, distribution-free estimators and tests for non-stationary autoregressive processes. *Stat Pap*. 2014;55(1):125-140. https://doi.org/10.1007/s00362-012-0479-5

40. Fertig KW, Meyer ME, Mann NR. On constructing prediction intervals for samples from a Weibull or extreme value distribution. *Technometrics*. 1980;22(4):567-573. https://doi.org/10.1080/00401706.1980.10486206

41. Engelhardt M, Bain LJ. On prediction limits for samples from a Weibull or extreme-value distribution. *Technometrics*. 1982;24(2):147-150. https://doi.org/10.1080/00401706.1982.10487738

42. Krishnamoorthy K, Lin Y, Xia Y. Confidence limits and prediction limits for a Weibull distribution based on the generalized variable approach. *J. Stat. Plann Inference*. 2009;139(8):2675-2684. https://doi.org/10.1016/j.jspi.2008.12.010

43. Frey J. Prediction bands for the EDF of a future sample. *J Stat Plann Inference*. 2012;142(2):506-515. https://doi.org/10.1016/j.jspi.2011.08.008

44. Xiong C, Milliken GA. Prediction for exponential lifetimes based on step-stress testing. *Commun Stat Simulation Comput*. 2002;31(4):539-556.

45. Hong Y, Meeker WQ. Field-failure predictions based on failure-time data with dynamic covariate information. *Technometrics*. 2013;55(2):135-149. https://doi.org/10.1080/00401706.2013.765324

46. Ross SM. *Introduction to Probability Models*. 11th. Amsterdam: Elsevier/Academic Press; 2014.

47. Baumgartner B, Gatto R. Value at ruin and tail value at ruin of the compound poisson process with diffusion and efficient computational methods. *Methodol Comput Appl Probab*. 2014;16(3):561-582. https://doi.org/10.1007/s11009-012-9316-5

48. Botev ZI, L'Ecuyer P, Rubino G, Simard R, Tuffin B. Static network reliability estimation via generalized splitting. *INFORMS J on Computing*. 2013;25(1):56-71. https://doi.org/10.1287/ijoc.1110.0493

49. Bates D, Maechler M. Matrix: Sparse and Dense Matrix Classes and Methods. R package version 1.2-4; 2016. https://CRAN.R-project.org/package=Matrix

50. Szugat S, Heinrich J, Maurer R, Müller CH. Prediction intervals for the failure time of prestressed concrete beams. *A Mater Scie Engine*. 2016;2016(9605450):8. https://doi.org/10.1155/2016/9605450

51. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc*. 2007;102(477):359-378. https://doi.org/10.1198/016214506000001437

52. Sánchez-Silva M, Klutke GA. *Reliability and Life-Cycle Analysis of Deteriorating Systems*. Heidelberg: Springer; 2016.

53. Lawless JF, Thiagarajah K. A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics*. 1996;38(2):131-138. https://doi.org/10.1080/00401706.1996.10484458

54. Dachian S, Kutoyants YA. Hypotheses testing: Poisson versus self-exciting. *Scand J Stat*. 2006;33(2):391-408. https://doi.org/10.1111/j.1467-9469.2005.00484.x

55. Berg MP, Haberman S. Trend analysis and prediction procedures for time nonhomogeneous claim processes. *Insur Math Econ*. 1994;14(1):19-32. https://doi.org/10.1016/0167-6687(94)00002-6

56. Franz J, Jokiel-Rokita A, Magiera R. Prediction in trend-renewal processes for repairable systems. *Stat Comput*. 2014;24(4):633-649. https://doi.org/10.1007/s11222-013-9393-5

57. Kopperschmidt K, Stute W. The statistical analysis of self exciting point processes. *Stat Sin*. 2013;23(3):1273-1298.

58. Chan NH, Li ZR, Yau CY. Forecasting online auctions via self-exciting point processes. *J Forecasting*. 2014;33(7):501-514. https://doi.org/10.1002/for.2313

59. Grothe O, Korniichuk V, Manner H. Modeling multivariate extreme events using self-exciting point processes. *J Econ*. 2014;182(2):269-289.

60. Hardiman SJ, Bouchaud JP. Branching-ratio approximation for the self-exciting Hawkes process. *Phys Rev E*. 2014;90:062807. https://doi.org/10.1103/PhysRevE.90.062807

61. Rambaldi M, Pennesi P, Lillo F. Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach. *Phys Rev E*. 2015;91:012819. https://doi.org/10.1103/PhysRevE.91.012819

62. Wheatley S, Filimonov V, Sornette D. The Hawkes process with renewal immigration and its estimation with an EM algorithm. *Comput Statis Data Analy*. 2016;94:120-135.

63. Kazemipour A, Babadi B, Wu M. Sparse estimation of self-exciting point processes with application to LGN neural modeling. In: 2014 IEEE Global Conference on Signal and Information Processing. GlobalSIP; 2014:478-482.

64. Bai F, Chen F, Chen K. Semiparametric estimation of a self-exciting regression model with an application in recurrent event data analysis. *Stat Sin*. 2015;25(4):1503-1526. https://doi.org/10.5705/ss.2013.217

65. Guo C, Luk W. Accelerating parameter estimation for multivariate self-exciting point processes. In: Proceedings of the 2014 ACM/SIGDA International Symposium on Field-programmable Gate Arrays. ACM; 2014:181-184.

66. Kustosz C, Szugat S. rexpar: Simplicial Depth for Explosive Autoregressive Processes. R package version 1.1. https://github.com/ChrisKust/rexpar; 2016.

AUTHOR BIOGRAPHIES

Kevin Leckey is a postdoctoral researcher at the Faculty of Statistics at TU Dortmund University. He received his PhD in mathematics at the Goethe University Frankfurt. His research interests include technometrics, robust statistics, and the probabilistic analysis of algorithms.

Christine Müller is a full professor of Statistics with Applications in Engineering Sciences at the Department of Statistics at TU Dortmund University. She received her PhD in mathematics at the Free University of Berlin. Before she became a professor in Dortmund, she was a lecturer in Göttingen and a professor in Oldenburg and Kassel. She is Coordinating Editor-in-Chief of Statistical Papers. From 2013 to 2019, she was the Chairwoman of the Deutsche Arbeitsgemeinschaft Statistik (DAGStat), an umbrella association of 14 German societies in statistics. Her research areas are technometrics, experimental design, and robust statistics

Sebastian Szugat received his master in Data Science at TU Dortmund University. He was a collaborator at the Chair of Statistics with Applications in Engineering Sciences at the Department of Statistics at the TU Dortmund University and is now working as a data scientist at Allianz Lebensversicherungs-AG in Stuttgart.

Reinhard Maurer is a full professor of Concrete Structures at the Department of Architecture and Civil Engineering at TU Dortmund University. He received his PhD in civil engineering at TH Darmstadt. Since 2001, he is working as a managing partner of the Design LLC König und Heunisch (KHP). He is an accredited checking engineer for structural design of solid construction and metal construction and an accredited checking engineer of the German Federal Railway Organisation for railway bridge building and civil engineering.

APPENDIX A: DERIVATIONS, PROOFS, AND IMPLEMENTATION DETAILS

A.1 | Derivation of the classical confidence sets

In order to also obtain an (asymptotic) confidence set for θ, we need additional assumptions for the design of the stress levels and total numbers of failures.

Let $\boldsymbol{d}_J := ((1, s_0), \dots, (I_0, s_0), \dots, (1, s_J), \dots, (I_J, s_J))$ be the concrete design. Moreover, let e_ω denote the Dirac measure in a point $\omega = (i, s) \in \Omega$, (i.e., $e_\omega(A) = 1$ if $\omega \in A$ and $e_\omega(A) = 0$ if $\omega \notin A$, where $A \subset \Omega$) and let

$$\delta_J = \frac{1}{N(J)} \sum_{j=0}^{J} \sum_{i=1}^{I_j} e_{(i,s_j)}, \qquad N(J) = \sum_{j=0}^{J} I_j$$

be the corresponding design measure on $\Omega := \{1, \dots, I\} \times [0, s_{max}]$ where s_{max} is the maximum possible stress level. Then, we assume $\delta_J \to \delta$ weakly as $J \to \infty$ for some design measure δ on $\{1, \dots, I\} \times [0, s_{max}]$. This implies[35, pp. 421-428]

$$\sqrt{N(J)} \, (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to \mathcal{N}\left(0, \boldsymbol{I}_\theta(\delta)^{-1}\right), \tag{A1}$$

where the asymptotic information matrix $\boldsymbol{I}_\theta(\delta)$ for exponentially distributed observations is given by[36]

$$\boldsymbol{I}_\theta(\delta) = \int \frac{1}{\lambda_\theta(i-1, s)^2} \, \dot{\lambda}_\theta(i-1, s) \, \dot{\lambda}_\theta(i-1, s)^\top \, \delta(d(i, s))$$

with $\dot{\lambda}_\theta(i, s) := \frac{\partial}{\partial \theta} \lambda_\theta(i, s)$. This matrix can be estimated by $1/N(J) \cdot \boldsymbol{I}(\widehat{\boldsymbol{\theta}})$ with $\boldsymbol{I}(\boldsymbol{\theta})$ given by (7). The continuous mapping theorem and Lemma of Slutzky provide that $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \, \boldsymbol{I}(\widehat{\boldsymbol{\theta}}) \, (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges weakly to a χ^2-distribution so that the set given by (6) is an asymptotic $(1 - \alpha)$-confidence set for $\boldsymbol{\theta}$. Using an additional approximation with the Taylor expansion for the i.i.d. case leads to the confidence sets based on the likelihood-ratio test given by (8).

Moreover, the δ-method applied to (A1) provides the asymptotic normality of $\sqrt{N(J)}(g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta}))$. Hence, an asymptotic $(1 - \alpha)$-confidence interval for $g(\boldsymbol{\theta})$ is given by (9).

A.2 | Proofs

Proof of Lemma 1. It is a well-known fact that $W/\lambda \sim \text{Exp}(\lambda)$ for $W \sim \text{Exp}(1)$ and $\lambda > 0$. Hence, we may assume that for all $\theta \in \Theta$,

$$W_n(\theta) = \frac{W_n}{\lambda_n(\theta)}, \quad n = 1, \dots, N, \tag{A2}$$

where $W_1, \dots, W_N$ are i.i.d. standard exponentially distributed random variables. Now, let $\theta^* \in \Theta$ be fixed. Since we consider independent exponentially distributed random variables, one can compute the log-likelihood function $\mathcal{L}$ for $(W_1(\theta^*), \dots, W_N(\theta^*))$ explicitly and obtains

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \log(\lambda_n(\theta)) - \sum_{n=1}^{N} \lambda_n(\theta) W_n(\theta^*), \quad \theta \in \Theta.$$

Using (5) and (A2), this may be rewritten as

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \log\left(\lambda_n(\theta^*)\right) + \sum_{n=1}^{N} \log\left(h_n(\theta - \theta^*)\right) - \sum_{n=1}^{N} h_n(\theta - \theta^*) W_n.$$

Hence, the maximum likelihood estimation (MLE) is given via $\hat{\theta}(\theta^*) = \theta^* + \widehat{\boldsymbol{\vartheta}}$, where $\widehat{\boldsymbol{\vartheta}}$ is the global maximum of the function

$$\widetilde{\mathcal{L}}(\widehat{\boldsymbol{\vartheta}}) = \sum_{n=1}^{N} \log\left(h_n(\widehat{\boldsymbol{\vartheta}})\right) - \sum_{n=1}^{N} h_n(\widehat{\boldsymbol{\vartheta}}) W_n.$$

Note that $\widehat{\boldsymbol{\vartheta}}$ does not depend on θ^*, that is, the distribution of the difference $\hat{\theta}(\theta^*) - \theta^*$ does not depend on the true model parameter θ^*. Consequently, once again using (A2),

$$\tilde{W}_n(\theta^*) = \lambda_n(\hat{\theta}(\theta^*)) \cdot \frac{W_n}{\lambda_n(\theta^*)} = h_n(\widehat{\boldsymbol{\vartheta}}) W_n.$$

The assertion follows since the right-hand side does not depend on θ^*.

Proof of Theorem 1. Since W_* and W are independent, we obtain for all θ_0

$$\lim_{J \to \infty} P_{\theta_0}(W \in \hat{I}_{pred}^1(W_*))$$

$$= \lim_{J \to \infty} P_{\theta_0}\left(W \in \bigcup_{\theta \in \hat{C}_{\alpha(1)}(W_*)} [b_{\alpha(2)/2}(\theta), b_{1-\alpha(2)/2}(\theta)] \right)$$

$$\geq \lim_{J \to \infty} P_{\theta_0}\left(W \in [b_{\alpha(2)/2}(\theta_0), b_{1-\alpha(2)/2}(\theta_0)] \quad \text{and} \quad \theta_0 \in \hat{C}_{\alpha(1)}(W_*) \right)$$

$$= \lim_{J \to \infty} P_{\theta_0}\left(W \in [b_{\alpha(2)/2}(\theta_0), b_{1-\alpha(2)/2}(\theta_0)] \right) P_{\theta_0}\left(\theta_0 \in \hat{C}_{\alpha(1)}(W_*) \right)$$

$$\geq (1 - \alpha(2))(1 - \alpha(1)).$$

□

Proof of Theorem 2. The one-sided confidence intervals given by (17) and (18) provide

$$P_\theta\left(b_{\alpha(2)/2}(\theta) < b_{\alpha(2)/2}(\widehat{\theta}) - \hat{v}_1 \right) \leq \frac{\alpha(1)}{2}$$

and

$$P_\theta\left(b_{1-\alpha(2)/2}(\theta) > b_{1-\alpha(2)/2}(\widehat{\theta}) + \hat{v}_2 \right) \leq \frac{\alpha(1)}{2}.$$

Since W and $\widehat{\theta} = \hat{\theta}(W_*)$ are independent, we obtain

$$\lim_{J \to \infty} P_\theta(W \in \hat{I}_{pred}^2(W_*))$$

$$= \lim_{J \to \infty} P_\theta\left(W \geq b_{\alpha(2)/2}(\widehat{\theta}) - \hat{v}_1 \quad \text{and} \quad W \leq b_{1-\alpha(2)/2}(\widehat{\theta}) + \hat{v}_2 \right)$$

$$\geq \lim_{J \to \infty} P_\theta\left(W \geq b_{\alpha(2)/2}(\theta) \quad \text{and} \quad b_{\alpha(2)/2}(\theta) \geq b_{\alpha(2)/2}(\widehat{\theta}) - \hat{v}_1 \quad \text{and} \right.$$

$$\left. W \leq b_{1-\alpha(2)/2}(\theta) \quad \text{and} \quad b_{1-\alpha(2)/2}(\theta) \leq b_{1-\alpha(2)/2}(\widehat{\theta}) + \hat{v}_2 \right)$$

$$= \lim_{J \to \infty} P_\theta\left(b_{\alpha(2)/2}(\theta) \leq W \leq b_{1-\alpha(2)/2}(\theta) \right)$$

$$\cdot \left(1 - P_\theta\left(b_{\alpha(2)/2}(\theta) < b_{\alpha(2)/2}(\widehat{\theta}) - \hat{v}_1 \quad \text{or} \quad b_{1-\alpha(2)/2}(\theta) > b_{1-\alpha(2)/2}(\widehat{\theta}) + \hat{v}_2 \right) \right)$$

$$\geq (1 - \alpha(2)) \cdot \left(1 - \frac{\alpha(1)}{2} - \frac{\alpha(1)}{2} \right) = (1 - \alpha(1))(1 - \alpha(2)).$$

□

A.3 | Implementation details

For details of the implementation in *R*, see Supporting Information. Here it should only be mentioned that some effort was necessary not only for the calculation of the quantiles of the hypoexponential distribution but also for the 3-depth method. In order to speed up the computation of the 3-depth, we use an (asymptotically) equivalent form[25, Equation (3.2)] that can be computed in linear time. The quantiles of the asymptotic distribution of the 3-depth were taken from the R package *rexpar*.[66] For small sample sizes, we used simulated quantiles of the distribution of the 3-depth calculated by Melanie Horn (TU Dortmund).

SUPPORTING INFORMATION

R-package loadshare:	R-package which provides the methods and data set. This package is available on GitHub and can be installed via `devtools::install_github(KLeckey/loadshare)`
More details and results:	R implementation details and more results of the simulation study can be found in the separate file `Loadshare_Supplementary_Material.pdf` given on https://www.statistik. tu-dortmund.de/leckey0.html.