International Journal of Fatigue 167 (2023) 107326

Contents lists available at ScienceDirect

International Journal of Fatigue

journal homepage: www.elsevier.com/locate/ijfatigue

Robust quantile regression analysis for probabilistic modelling of *S-N* curves

Qingrong Zou [a], Jianxi Zhao [a], Jici Wen [b,c,*]

[a] School of Applied Science, Beijing Information Science and Technology University, Beijing 100192, China
[b] State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Engineering Sciences, University of Chinese Academy of Sciences, China

ARTICLE INFO

Keywords:
S-N curves
Fatigue life
Quantile regression
Fatigue design

ABSTRACT

The scatter in fatigue data is commonly characterized by probability distributions for constructing the probabilistic *S-N* curves. However, there is notable estimation bias under distribution misspecification. In this paper, we proposed a quantile regression framework for modeling *S-N* curves. The quantile regression model can be built directly on the experimental data without any distribution assumption. Extensive simulations and two experimental datasets are used to illustrate the usefulness of the proposed model. The results demonstrate that the quantile regression model is exempt from the problem of incorrectly specifying the potential fatigue life distribution and is robust to the non-constant scale problem.

1. Introduction

Engineering systems often fatigue and fail catastrophically when they are subjected to repeated stress over a long time [12,29,44,37]. Fatigue life, defined as the number of stress cycles under cyclic loading [15], is an important indicator to evaluate the fatigue rule, and then applied for reliability design in engineering [24,21,6,40,36]. The *S-N* curve is an important criterion to depict the relationship between cyclic stress and fatigue life [20,5,13]. However, fatigue in materials mainly originates from microscopic material flaws to grow into a macroscopic crack under cyclic stress conditions, so the fatigue test results usually exhibit large dispersion due to the fluctuation of microscopic imperfections in distributions and sizes [12,30,34]. Therefore, it is virtual important for evaluating the dispersion of fatigue life by extending the median *S-N* curve into the p quantile *S-N* curves, namely probabilistic *S-N* (*P-S-N*) curves, a generalization that relates the p quantile of fatigue life to the applied stress [6].

The variability in stress-life datasets is usually described by probability distribution to enable the construction of the *P-S-N* curves for design. Some early research assumed a normal distribution for fatigue life [9,31]. Most of the previous studies in probability *S-N* assumed that the fatigue life follows a log location scale distribution, where the lognormal and Weibull distributions are the two most commonly used distributions for the analysis of the probabilistic behavior of fatigue failure [28].

For example, ISO 12107:2003 assumed that the fatigue life followed a lognormal distribution to derive the *P-S-N* curves [15]. However, the variances of fatigue lives at all stress levels are considered the same. Xie et al. converted all data on fatigue into equivalent stress values and the lognormal distribution was used for *P-S-N* curves [39]. A hierarchical Bayesian model integrated with Hierarchical Bayesian data augmentation was proposed to deal with sparse data problem for fatigue *S-N* curves, where the fatigue life was assumed to follow a lognormal distribution with heteroscedasticity at different stress levels [6]. Chen and Liu proposed a probabilistic physic-guided neural network for *P-S-N* curve estimation by modeling both the mean and variance of fatigue life with a lognormal assumption [7]. Zu et al. proposed a novel *S-N* curve modeling method with uncertainty theory, and the logarithm failure life was assumed to follow normal uncertainty distribution [42].

Since the Weibull distribution was put forward in 1939 [35], 2-parameter or 3-parameter Weibull distributions have been frequently used to analyze fatigue life data [41,4,33,17,14]. For instance, Júnior and Belísio associated the Weibull probability equations with the usually applied formula when constructing the *S-N* curves for composite materials, which are the power law and the exponential equation and their generalizations [16]. The lognormal distribution, the 2-parameter Weibull, the Gumbel, and the 3-parameter Weibull modes were adopted by Mohabeddine et al. to assess the fatigue behavior of fiber reinforced polymer retrofitted specimens [22].

It is worth noting that a major limitation of most existing stress-life models is that they have assumed a specific distribution for fatigue life data. However, life quantiles, which are of interest to engineering

* Corresponding author at: State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: wenjici@lnm.imech.ac.cn (J. Wen).

https://doi.org/10.1016/j.ijfatigue.2022.107326
Received 3 August 2022; Received in revised form 7 October 2022; Accepted 8 October 2022
Available online 15 October 2022

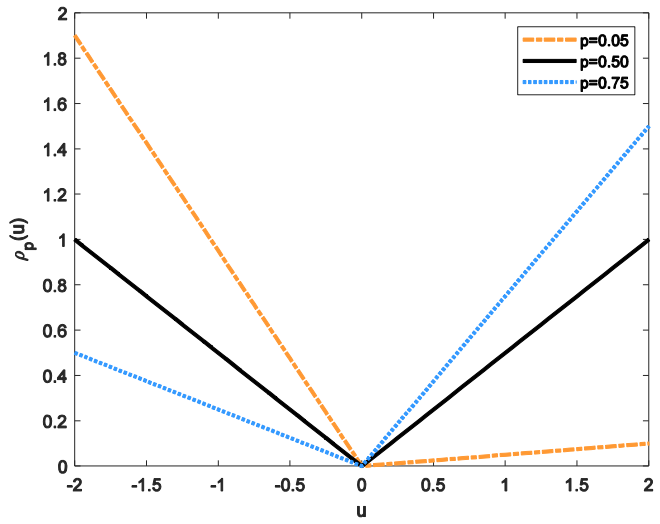

Fig. 1. Illustration of the check function.

Table 1
Summary of simulation scenarios.

Distribution of fatigue life N	Location parameter	Scale parameter $\sigma(x)$
Lognormal	$\alpha_0 = 12,\ \alpha_1 = -3$	$\sigma(x) = 0.5$
Lognormal	$\alpha_0 = 12,\ \alpha_1 = -3$	$\sigma(x) = 0.5 - 0.15x$
Weibull	$\alpha_0 = 12,\ \alpha_1 = -3$	$\sigma(x) = 0.5$
Weibull	$\alpha_0 = 12,\ \alpha_1 = -3$	$\sigma(x) = 0.5 - 0.15x$
Log-Logistic	$\alpha_0 = 12,\ \alpha_1 = -3$	$\sigma(x) = 0.5$
Log-Logistic	$\alpha_0 = 12,\ \alpha_1 = -3$	$\sigma(x) = 0.5 - 0.15x$

applications, may differ significantly between a hypothetical distribution and the actual distribution [25]. In addition, most previous works considered a constant dispersion parameter at all stress levels, which has been found inappropriate in many applications, for example, metal fatigue and many electronic wear-out failures.

Motivated by the fact that a lower quantile of the fatigue life distribution under service conditions is of particular interest in fatigue problems, this study proposes a quantile regression (QR) framework for fatigue data analysis. Firstly, we establish an QR model for *P-S-N* curves, which can effectively estimate the conditional quantile of the distribution of fatigue life, and no specific distribution assumption for the fatigue life data is required. This means that our approach is robust to the underlying distributions. Secondly, we readily dealt with the heteroscedasticity in the framework via allowing the dispersion parameter to be a function of the stress. Thirdly, multiple probabilistic *S-N* curves can be simultaneously obtained by stepwise quantile regression estimation with non-crossing constraints.

2. Preliminaries on quantile regression

Regression is applied to quantify the association between a dependent variable and some independent variables. The QR approach proposed by Koenker and Basset [18] is a vital progress in regression analysis, which is more appropriate for asymmetric and long-tailed distributions or when the tails of the underlying distributions are of interest for modeling extreme behavior [18]. A great number of research and applications indicate that the QR has a lot of advantages. One of the most attractive merits is its ability to estimate full conditional quantile effects that depict the impact of independent variables not only on the median but also on the tails of the distribution of the dependent variable.

Let y_i denotes the i-th observation of the dependent variable and x_i represents the corresponding independent variable vector. The linear QR model with p-th quantile $(0 < p < 1)$ for the dependent variable y_i given x_i as.

$$y_i = x_i^T \beta_p + \varepsilon_i,\ \ i = 1, 2, \cdots, n, \tag{1}$$

where $\boldsymbol{\beta_p}$ is the regression coefficient vector, ε_i is the error term whose p-th quantile is assumed to be zero.

The conditional median and other quantile functions can be estimated by minimizing asymmetrically weighted absolute residuals. That is, $\boldsymbol{\beta_p}$ can be estimated by minimizing the following loss function:

$$\frac{1}{n}\sum_{i=1}^{n} \rho_p(y_i - x_i^T \boldsymbol{\beta_p}) \ . \tag{2}$$

where $\rho_p(\cdot)$ is the check (or loss) function, which satisfies $\rho_p(u) = pu$ if $u > 0$ and $\rho_p(u) = (p-1)u$ if$u \leqslant 0$. An illustration of the check function with several distinct values of p is shown in Fig. 1. Although the check function is not differential at the points with zero residuals, it has directional derivative in all directions. Therefore, the standard linear programming can be used for minimizing Eq. (2).

3. Methodology

For n specimens, they are performed in the fatigue life test at m stress levels, in which$S_1 < S_2 < \cdots < S_m$, and there are n_i fatigue data obtained at the i-th stress level. Then the fatigue data is$\{(S_i, N_{ij}), i = 1, 2, \cdots, m, j = 1, 2, \cdots n_i\}$, where N_{ij} expresses the fatigue life of the j-th specimen at stress level S_i and$\sum_{i=1}^{m} n_i = n$..

The fatigue properties of products or materials are determined by testing a set of specimens over different stress levels to construct a fatigue life relationship as a function of stress. In engineering applications, a linear stress-life curve (on a log–log scale) is generally assumed [43], i.e., the Basquin relation [3]:

$$N_{ij} = AS_i^{-B}, \tag{3}$$

where A > 0 and B > 0 are the fatigue curve coefficients which are shared across stress levels.

The linear expression is obtained by taking the logarithm transformation of Eq. (3),

$$\log N_{ij} = \log A - B \log S_i \ . \tag{4}$$

Denote$x_i = \log S_i,\ y_{ij} = \log N_{ij},\ \alpha_0 = \log A,\ \alpha_1 = -B$, Eq.(4) can be rewritten as.

$$y_{ij} = \alpha_0 + \alpha_1 x_i \ . \tag{5}$$

Scattering is inherent to fatigue data due to several sources related to the experimental variability and the discrepancy of product quality. Hence, the fatigue life is random in nature. Consider the logarithm fatigue life y and the logarithm stress level x following the general relationship.

$$y_{ij} = \alpha_0 + \alpha_1 x_i + \sigma(x_i) \cdot \varepsilon_{ij} \ , \tag{6}$$

where $\sigma(x_i) > 0$ is a deterministic scale function, and ε_{ij} is an arbitrary random variable which determines the distribution ofy_{ij}. Next, the quantile regression is firstly constructed under the constant assumption, namely,$\sigma \equiv \sigma(x_i) > 0$, then we extend it to the case with non-constant$\sigma(x_i)$..

3.1. Derivation of P-S-N curves with constant scale function

We firstly assume the scale σ is a constant. Let $Q_{y_{ij}|x_i}(p)$ and $Q_{\varepsilon_{ij}}(p)$ denote the p quantile of y_{ij} and ε_{ij} givenx_i, respectively. Based on Eq. (6) and the equivariance property [19], we have.

$$Q_{y_{ij}|x_i}(p) = \alpha_0 + \alpha_1 x_i + \sigma \cdot Q_{\varepsilon_{ij}}(p) \ . \tag{7}$$

In contrast to regression model that focuses on$E(y_{ij}) = \alpha_0 + \alpha_1 x_i$, quantile regression aims to describe the quantile relationship shown in

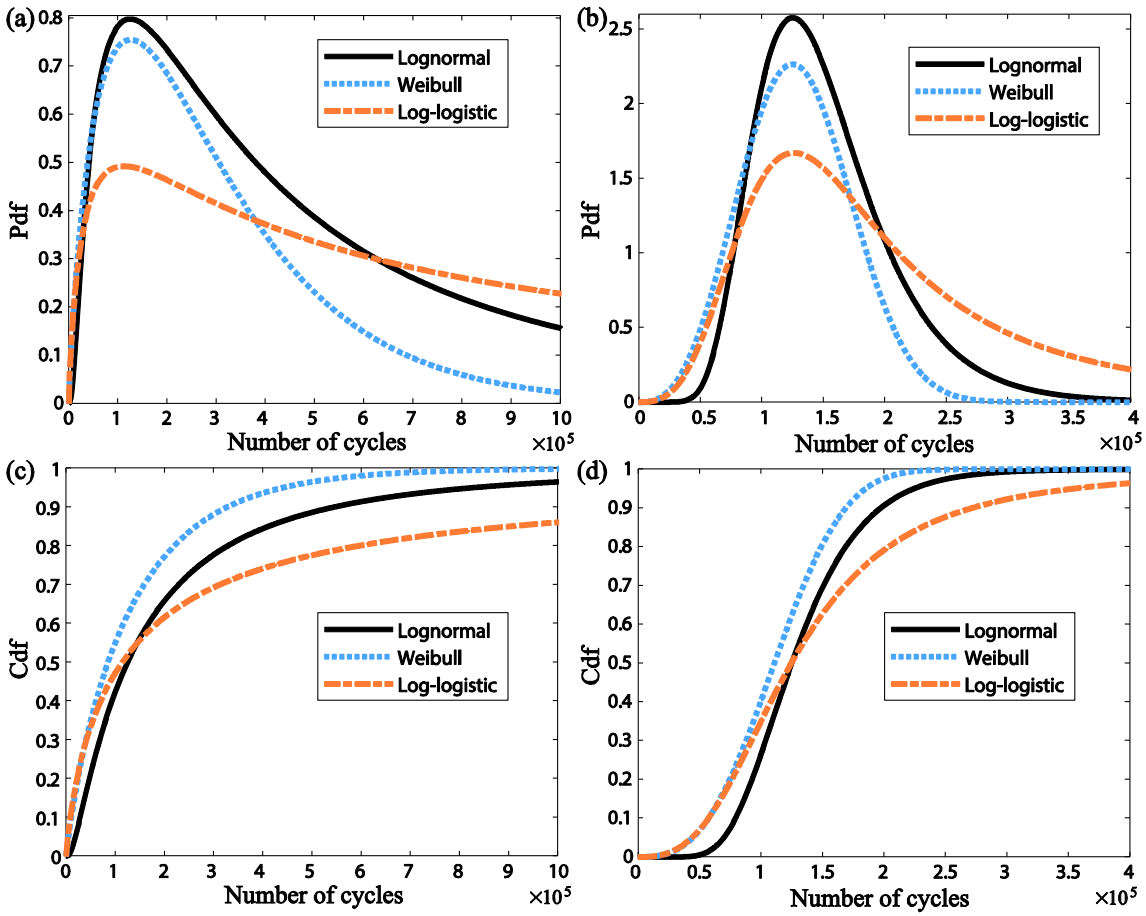

Fig. 2. Probabilistic characteristics of three distributions at the stress level 200 MPa. (a & b) The probability density function with different scale parameters: (a) σ = 0.5, (b) σ = 0.5–0.15x; (c & d) The cumulative distribution functions with different scaling parameters: (c) σ = 0.5, (d) σ = 0.5–0.15x.

Eq. (7). The Eq. (7) can be rewritten as.

$$Q_{y_{ij}|x_i}(p) = \beta_0(p) + \beta_1(p)x_i \ , \tag{8}$$

where $\beta_0(p) = \alpha_0 + \sigma Q_{\varepsilon_{ij}}(p)$ and $\beta_1(p) = \alpha_1$.

Given the logarithm fatigue data $\{(x_i, y_{ij}), i = 1,2,\cdots,m, j = 1,2,\cdots n_i\}$, the parameter vector $\boldsymbol{\theta}(p) = (\beta_0(p), \beta_1(p))^T$ can be estimated as.

$$\widehat{\boldsymbol{\theta}}_n(p) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \rho_p(y_{ij} - \beta_0(p) - \beta_1(p)x_i) \ . \tag{9}$$

With the estimated parameter vector $\widehat{\boldsymbol{\theta}}(p)$, the p quantile of the logarithm failure life at the applied logarithm stress x can be predicted as $\widehat{Q}_{y|x}(p) = \mathbf{x}^T \widehat{\theta}_n(p)$, where $\mathbf{x} = (1, x)^T$. Under some regularity conditions which can be found in [19], $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n(p) - \boldsymbol{\theta}(p))$ converges to a normal distribution with zero mean and covariance matrix.

$$\sum_{\theta(p)} = \lim_{n\to\infty} \frac{n\sigma^2 p(1-p)}{f^2(F^{-1}(p))} \left[\sum_{i=1}^{m} n_i \mathbf{x_i} \mathbf{x_i^T} \right]^{-1} , \tag{10}$$

where $\mathbf{x_i} = (1, x_i)^T$, $f(\cdot)$ and $F(\cdot)$ are the probability density function and cumulative distribution function of random variable ε, respectively. According to the theory for convergence of transformed random sequences, the forecasted quantile $\widehat{Q}_{y|x}(p)$ is consistent and asymptotically normal [19]. With the delta method, its asymptotically variance is as follows.

$$Var(\widehat{Q}_{y|x}(p)) = \frac{1}{n} \mathbf{x_0^T} \sum_{\theta(p)} \mathbf{x_0} \ . \tag{11}$$

3.2. Derivation of P-S-N curves with non-constant scale function

When the stress not only impacts the location parameter but also impacts the scale parameter of the distribution of y, unequal scale function $\sigma(x)$ is required. In addition, for the majority of engineering materials, the standard deviation of fatigue life commonly decreases with increased cyclic stress levels [8,32,10]. Therefore, we assume $\sigma(x)$ is a linear function of x, i.e., $\sigma(x) = a_0 + a_1 x$. According to Eq. (6), the p quantile of y_{ij} can be written as.

$$Q_{y_{ij}|x_i}(p) = \beta_0(p) + \beta_1(p)x_i \ , \tag{12}$$

where $\beta_0(p) = \alpha_0 + a_0 Q_{\varepsilon_{ij}}(p)$ and $\beta_1(p) = \alpha_1 + a_1 Q_\varepsilon(p)$. After this reformulated, we can see that Eq. (12) is same as the Eq. (8), which is appealing since we do not need to judge whether the scale parameter is constant in advance. Hence, the parameter vector $\boldsymbol{\theta}(p)$ is also estimated by Eq. (9). Like the constant scale case, the estimator $\widehat{\boldsymbol{\theta}}_n(p)$ is consistent and asymptotic normal. However, the asymptotic covariance matrix shown in Eq. (13) differs slightly from Eq. (10) because of the heterogeneity.

$$\sum_{\theta(p)} = \lim_{n\to\infty} \frac{n\sigma^2 p(1-p)}{f^2(F^{-1}(p))} \left[\sum_{i=1}^{m} \frac{n_i \mathbf{x_i} \mathbf{x_i^T}}{a_0 + a_1 x_i} \right]^{-1} \left[\sum_{i=1}^{m} n_i \mathbf{x_i} \mathbf{x_i^T} \right] \left[\sum_{i=1}^{m} \frac{n_i \mathbf{x_i} \mathbf{x_i^T}}{a_0 + a_1 x_i} \right]^{-1} . \tag{13}$$

In practical application, the explicit solution of Eq. (10) and Eq. (13) may not be available due to the unknown distribution of ε. As an alternative, we can employ the bootstrap to compute the standard deviation for $\widehat{\boldsymbol{\theta}}_n(p)$ and also no need to know whether σ depends on the stress or not.

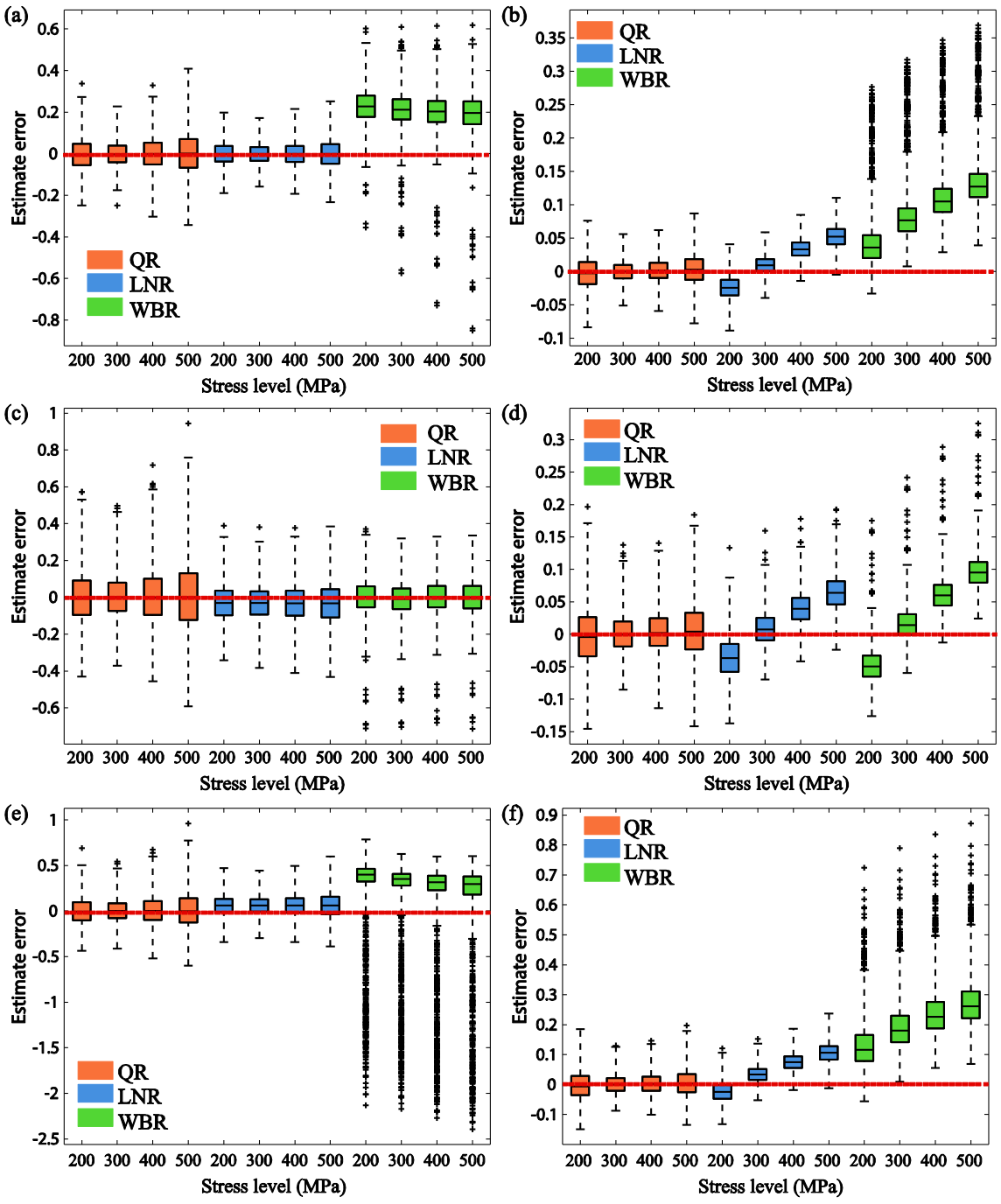

Fig. 3. Boxplot of the prediction errors at 10% failure probability under four applied stress levels. 50 samples are drawn at each stress level. Two different scale parameter scenarios are considered: (a, c, e) the homogeneous scenario, (b, d, f) the heteroscedasticity scenario. And three different underlying distributions: (a, b) the lognormal distribution; (c, d) the Weibull distribution, (e, f) the log-logistic distribution.

3.3. Non-crossing and monotonicity constraints

According to definition, for any given fixed values of the stress levels, conditional quantile regression functions are increasing functions in terms of quantiles p. However, separately estimated QR functions commonly cross each other, especially when the samples are small to moderate. It is annoying and undesirable for further utilization and interpretation of theses QR functions. In addition, each survival probability *S-N* curve should be non-increasing over stress (strain) levels. Motivated by Wu and Liu [38], we incorporated the above two points into QR models by imposing several linear inequality constraints to obtain reliable estimates of quantiles. For the convenience of expression, denote p_{k_0} as the start quantile, and the shorthand notation $\beta_{0,k}$ for$\beta_0(p_k)$, $\beta_{1,k}$ for$\beta_1(p_k)$, respectively. The detailed algorithm as follows:

Step 1: Initialize: While any percentile could be chosen as the starting point, some references [38,23] suggest using the p = 0.5 as the starting quantile since usually it has the lowest variance, therefore, the estimated quantile function is relatively more accurate. Here, we also start from the middle. The linear programming based on Barrodale and Roberts algorithm [1] can be applied to obtain parameter estimates $\widehat{\beta}_{0,k_0}$ and$\widehat{\beta}_{1,k_0}$.

Step 2: For $k = k_0 + 1, \cdots, k_0 + K$ ('higher' quantiles), we consider the estimation of $\beta_{0,k+1}$ and $\beta_{1,k+1}$ based on $\widehat{\beta}_{0,k}$ and $\widehat{\beta}_{1,k}$ by solving.

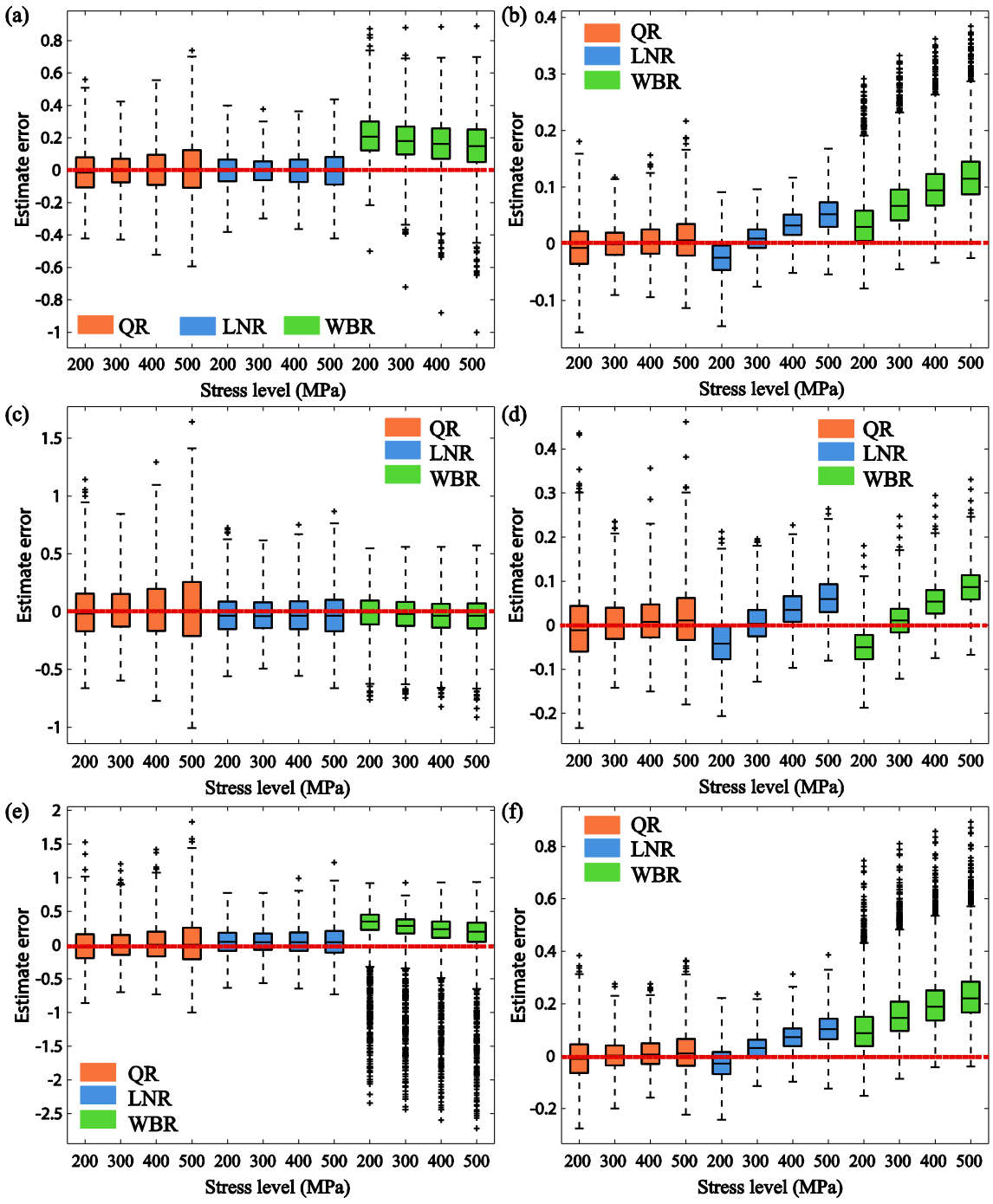

Fig. 4. Boxplot of the prediction errors at 10% failure probability and 15 samples are drawn at each stress level. Two different scale parameter scenarios are considered: (a, c, e) the homogeneous scenario, (b, d, f) the heteroscedasticity scenario. And three different distributions are given as: (a, b) the lognormal distribution; (c, d) the Weibull distribution, (e, f) the log-logistic distribution.

$$\min_{\beta_{0,k+1},\beta_{1,k+1}} \sum_{i=1}^{n} \rho_{p_{k+1}}(y_i - \beta_{0,k+1} - \beta_{1,k+1}x_i)$$

$$s.t. \begin{cases} \beta_{0,k+1} > 0, \beta_{1,k+1} < 0, \\ \beta_{0,k+1} + \beta_{1,k+1}x_{\min} > \widehat{\beta}_{0,k} + \widehat{\beta}_{1,k}x_{\min}, \\ \beta_{0,k+1} + \beta_{1,k+1}x_{\max} > \widehat{\beta}_{0,k} + \widehat{\beta}_{1,k}x_{\max}. \end{cases} \tag{14}$$

Step 3: For $k = k_0 - 1, \cdots, k_0 - K$ ('lower' quantiles), we consider the estimation of $\beta_{0,k-1}$ and $\beta_{1,k-1}$ based on $\widehat{\beta}_{0,k}$ and $\widehat{\beta}_{1,k}$ by solving.

$$\min_{\beta_{0,k-1},\beta_{1,k-1}} \sum_{i=1}^{n} \rho_{p_{k-1}}(y_i - \beta_{0,k-1} - \beta_{1,k-1}x_i)$$

$$s.t. \begin{cases} \beta_{0,k-1} > 0, \beta_{1,k-1} < 0, \\ \beta_{0,k-1} + \beta_{1,k-1}x_{\min} < \widehat{\beta}_{0,k} + \widehat{\beta}_{1,k}x_{\min}, \\ \beta_{0,k-1} + \beta_{1,k-1}x_{\max} < \widehat{\beta}_{0,k} + \widehat{\beta}_{1,k}x_{\max}. \end{cases} \tag{15}$$

Table 2
Fatigue life data of aluminum alloy 2524-T4 [27].

	Stress level (MPa)					
	371.7	411.7	451.7	490.3	530.3	550.3
No.	Cycles life to failure given stress					
1	213,100	131,800	78,400	51,100	31,700	27,200
2	218,900	133,900	79,200	53,200	32,400	27,600
3	222,500	136,700	79,800	53,500	33,200	28,900
4	234,000	137,100	82,300	54,200	33,700	29,300
5	241,500	139,300	85,800	55,000	35,500	29,900
6	246,800	140,100	86,000	57,000	35,700	30,100
7	248,600	140,400	86,600	57,000	35,900	30,600
8	251,200	145,700	88,100	57,200	36,000	30,700
9	251,300	145,900	88,500	57,700	36,500	30,800
10	252,700	146,000	88,700	58,400	36,800	30,900
11	255,600	146,900	89,300	58,500	37,000	30,900
12	261,600	148,600	90,300	59,200	37,000	31,000
13	261,600	148,600	90,800	59,300	37,200	31,200
14	262,400	148,800	92,200	59,800	37,500	31,200
15	262,900	149,400	92,500	59,800	37,500	31,700
16	265,300	152,500	93,200	60,900	37,600	31,800
17	265,700	153,600	93,400	61,400	37,600	31,900
18	268,700	154,200	93,700	61,400	37,900	32,100
19	274,000	155,000	94,100	61,600	38,200	32,400
20	280,500	155,800	95,300	62,500	38,600	32,400
21	282,100	156,200	95,500	62,900	39,400	34,900
22	284,800	156,700	95,600	63,000	39,600	
23	287,900	157,100	96,300	63,100	40,300	
24	289,300	159,200	96,600	63,300	40,500	
25	306,200	160,900	96,700	63,400	40,700	
26	310,100	161,000	99,200	63,600	40,800	
27	323,600	163,900	99,500	63,900	40,900	
28	326,200	164,300	107,500	66,700	41,000	
29	340,200	168,400	107,700	67,700	41,300	
30	343,200	183,200	110,600	70,000	42,700	

4. Simulated data experiments

In this section, the performance of the QR is compared with some existing parametric regression approaches, such as the regression model with Weibull distribution assumption (WBR) and the regression model with the lognormal distribution assumption(LNR), by using several simulation studies based on the Eq. (6). Three different distributions (Lognormal, Weibull, and Log-logistic) are simulated to demonstrate the property of distribution-free, and both the constant ($\sigma = 0.5$) and non-constant ($\sigma =$ 0.5–0.15x) variances are considered for demonstrating the robustness of the proposed model to heteroscedasticity. The different scenarios, a total of $3x2 = 6$ combinations of simulation conditions, are summarized in Table 1. The visual simulation distributions at the stress level 200 MPa are shown in Fig. 2. The distributions of the probability density function (Pdf) are shown with considering the homogeneous scenario in Fig. 2a and the heteroscedasticity scenario Fig. 2b, respectively. Fig. 2c and d are the distributions of the cumulative distribution function (Cdf) with considering the homogeneous scenario and the heteroscedasticity scenario, respectively.

For each simulation model, we firstly simulate 50 samples at 4 stress levels which are 150 MPa, 250 MPa, 350 MPa, and 450 MPa, respectively. The fatigue life and stress levels are logarithms transformed with

Table 4
List of experimental durability data for the S420MC steel [45].

S_j	n_j	N_j	S_j	n_j	N_j
204	1	946,200	232	6	488400, 380500, 567000, 701800, 553000, 630,000
207	1	1,851,500	248	6	286700, 376900, 488300, 650100, 585900, 698,500
210	1	1,281,700	250	6	313700, 256900, 238800, 323500, 213700, 389,000
211	1	1,215,000	267	3	199400, 194000, 224,800
214	8	628100, 1307600, 1316000, 1410600, 851900, 1566600, 959900, 1,159,400	268	6	120800, 139800, 159100, 187100, 219600, 238,600
219	2	1095800, 1,499,200	271	3	259500, 313000, 346,100
221	1	1,926,800	286	9	61600, 119400, 81600, 132000, 130000, 104300, 97400, 175600, 136,500
224	2	1999500, 997,600	295	3	136800, 129900, 151,400
229	6	690600, 730500, 1009600, 1555800, 1358000, 1,447,200			

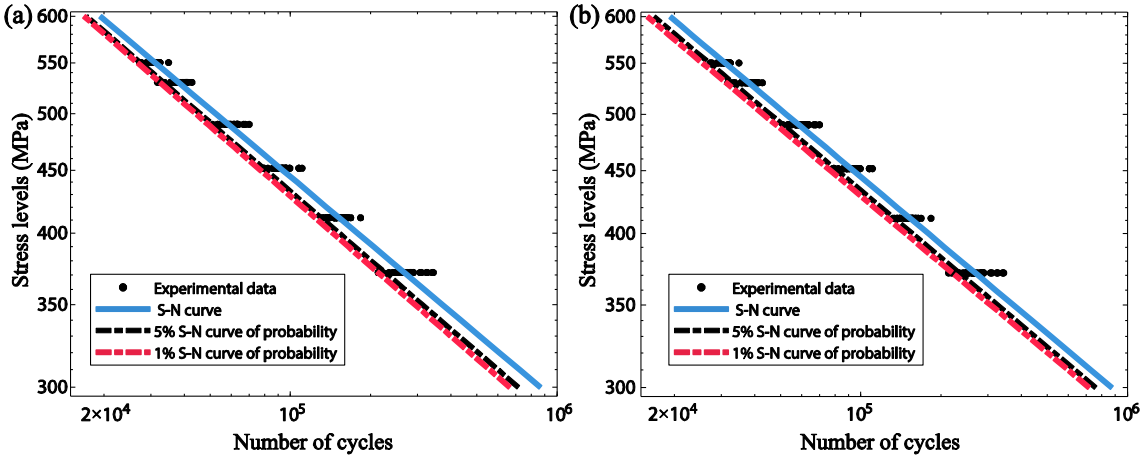

Fig. 5. Predictability to the probability *S-N* curves derived by two different models in log–log coordinate. (a) QR model, (b) ISO method.

Table 3
Variations of logarithm fatigue life.

Stress levels	371.7	411.7	451.7	490.3	530.3	550.3
Variations	2.84e−3	1.01e−3	1.35e−3	1.01e−3	1.04e−3	0.58e−3

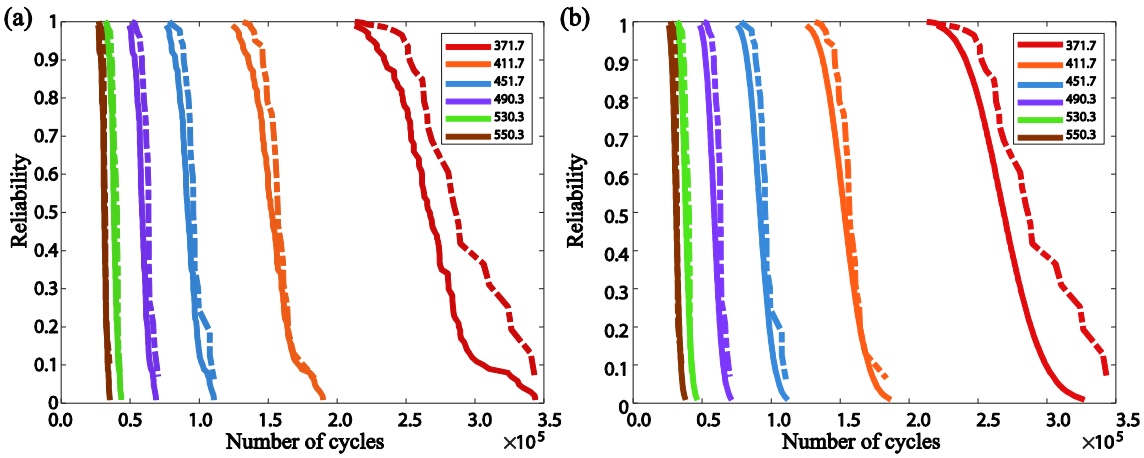

Fig. 6. Reliability of the 2024-T4 aluminum alloy at different stress levels. (a) Reliability results from the QR model (solid lines) and the K-M model (dotted lines). (b) Reliability results from the ISO model (solid lines) and the K-M model (dotted lines).

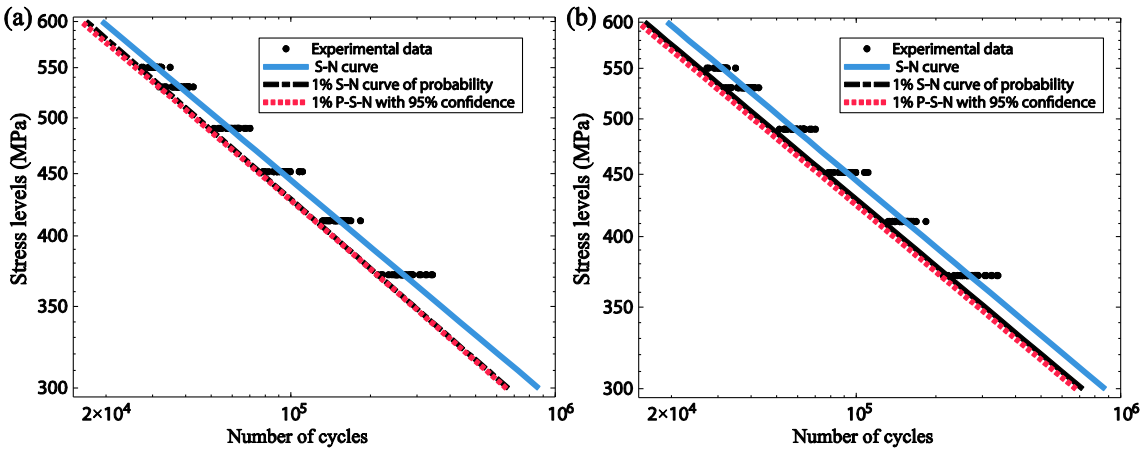

Fig. 7. 99% probability of survival *P-S-N* curve with 95% confidence tolerance bound with two different models in log–log coordinate. (a) QR model, (b)ISO method.

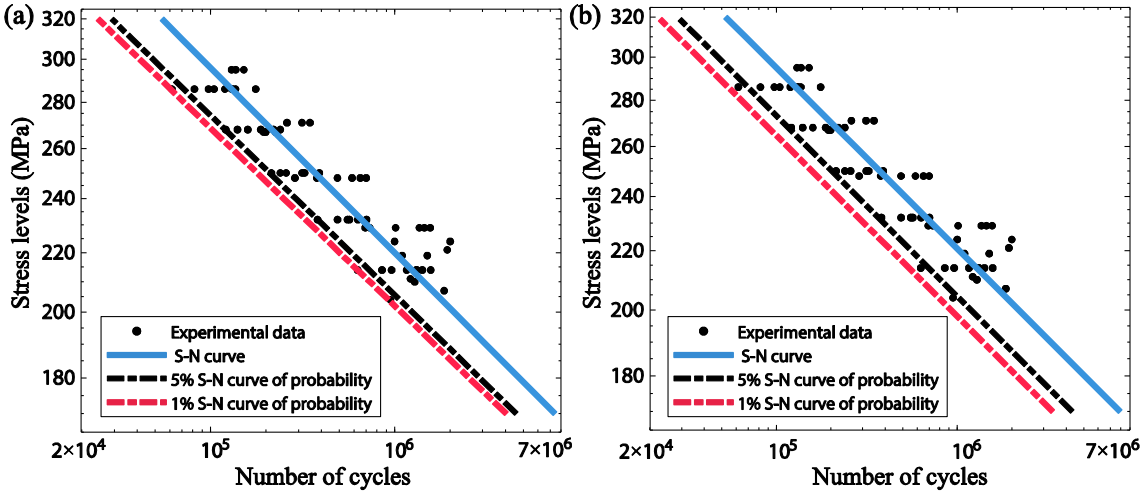

Fig. 8. *P-S-N* curves derived by the QR and ISO in log–log coordinate. (a) QR, (b) ISO.

base 10. For the LNR model, parameters α_0 and α_1 are estimated by least square estimation, and the scale parameter σ is estimated based on the sum of squares error [15]. For the WBR model, the parameters are estimated by maximum likelihood. It should be noted that both above two regression models assume a constant parameter over all stress levels. For the proposed QR model, conditional quantile functions are estimated from 0.01 to 0.99, step by 0.01. The performance of QR, LNR, and WBR are compared by 5000 repetitions.

The 0.10 quantile and 0.05 quantile, that is 90 % and 95 % survival probability, under four stress levels (200 MPa, 300 MPa, 400 MPa, and 500 MPa) are obtained from the estimated models. To evaluate model performance, prediction errors between the true quantile value and

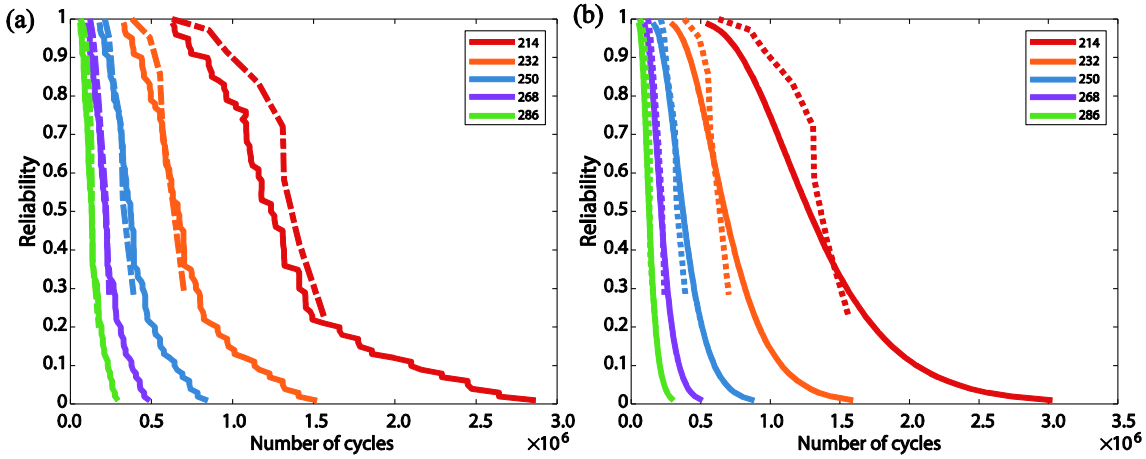

Fig. 9. Reliability of the S420MC steel at different stress levels (a) QR reflected by the solid line, and dash dot denotes K-M. (b)Solid lines are ISO and dash dot are K-M.

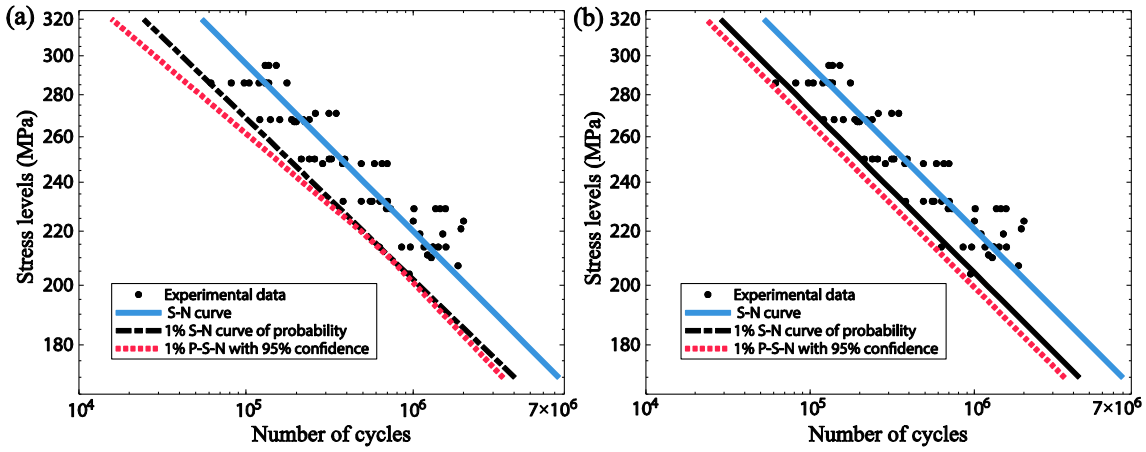

Fig. 10. 99% probability of survival *P-S-N* curve with 95% confidence tolerance bound in log–log coordinate. (a)QR, (b)ISO.

estimated quantile value are computed. Fig. 3 displays the boxplot of the prediction errors at 10 % failure probability under distinct simulation scenarios. Fig. 3a, c, and e show the prediction errors with three underlying distributions (Lognormal, Weibull, and log-logistic distributions in order) at the same constant scale parameter, i.e., σ is 0.5, and Fig. 3b, d, and f show the results at the non-constant scale parameter σ = 0.5–0.15x. The boxes symmetric with respect to the zero value mean the estimations are unbiased. In the homogeneous variance simulation scenario, the LNR model and the WBR model are unbiased only when the underlying distribution is lognormal and Weibull, respectively. In contrast, the proposed QR model generally obtains unbiased estimations under all underlying distributions. In the heteroscedasticity simulation scenario, we can see that the proposed QR model still performs satisfactorily under all underlying distributions. However, the prediction errors of the two parametric models are all seriously biased. We further explore the robustness of the QR model for extreme value behavior, such as 5 % and 2.5 % failure probability. The boxplots of the prediction errors at 5 % and 2.5 % failure probability under distinct simulation scenarios are shown in Figs. A1 and A2. We can see that the QR model also shows satisfactory performance in comparison with the other two models.

Due to the fatigue test usually being time-consuming and costly, it is difficult to carry out plenty of experiments. Therefore, we also simulate the scenario where only 15 samples are drawn for each stress level, and the estimation errors at 10 % failure probability are shown in Fig. 4. Similarly, the QR model gives the most satisfactory and almost unbiased estimation results for any underlying distributions. The predictability of the QR model at the two extreme failure probability conditions (5 % and 2.5 %), which is commonly concerned in engineering, is then checked and shown in Figs. A3 and A4. The above simulation results clearly reveal that the QR model is more appropriate to robustly estimate the *P-S-N* when the underlying distribution of failure life is unknown. In addition, it is also robust to heterogeneous variance.

5. Verifications

In order to further verify the performance of the QR model, we compared it with the ISO method [15] on two widely used real datasets, aluminum alloy 2024-T4 [27] and S420MC steel [45] in this section. The detailed experimental data are listed in Table 2 and Table 4.

5.1. 2024-T4 aluminum alloy data

The dataset of the fatigue life of aluminum alloy 2524-T4, as listed in Table 2, totally contains 171 experimental data points with 6 applied stress levels. Fig. 5a and b show the predicted probability *S-N* curve from the QR model and the ISO method, respectively. The median *S-N* curve (blue line in Fig. 5) for depicting the relationship between stress and cycle life, which means 50 % failure probability *S-N* curve, obeys well the Basquin relation in Eq. (3). We also show 5 % (black dotted line) and 1 % (red dotted line) failure probability *S-N* curve in Fig. 5. They both show the linear relationship in log–log coordinate predicted from the QR model in Fig. 5a and the ISO method in Fig. 5b. We observed in Fig. 5a a slight difference of the slope of the *S-N* curves at different failure probabilities. Namely, the QR model reveals the stress level-related

distribution properties from the experimental data. Restricted by the same underlying distribution assumption, the slopes of the *S-N* curves at different failure probabilities in the ISO method are the same.

To verify the stress level-related distribution property, we statistic the variations of the logarithm fatigue life at each stress level in Table 3. We can see that the variance of the logarithmic life at the highest stress level is far lower than that at the lowest stress level. To further determine whether or not the variances between serval stress levels are equal, we applied Bartlett's Test [2] based on a chi-square statistic. In particular, we wish to test the following null hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 \cdots = \sigma_k^2$$

against the alternative that variances are not equal for at least two stress levels, where k is the number of stress levels. Here, the Bartlett's chi-square value is 18.793, and the p-value is 0.0021 far lower than the significance level of 0.01. This means we reject the null hypothesis with sufficient evidence and conclude that the variances across stress levels are heterogeneous. So the performance of the QR model is more reasonable.

The Kaplan-Meier (K-M) estimator, also known as the product limit reliability estimator, is the most widely applied method to estimate the reliability function (the survival function) of the product life without any assumptions on the underlying distribution. Hence, the K-M estimator is applied for the fatigue reliability evaluation. As shown in Fig. 6, the reliability of the 2024-T4 aluminum alloy at different stress levels based on the QR model and ISO are compared with the K-M estimators. Firstly, compared to the K-M, both the QR model and ISO are conservative. Secondly, the reliability of the 2024-T4 aluminum alloy gradually decreased with the service cycles, which means that the fatigue failure of the 2024-T4 aluminum alloy increases gradually as the service time increases. In addition, the higher the stress level, the higher decline rate of the fatigue reliability curve.

When fatigue failure data are limited, the estimation of probabilistic *S-N* curve becomes uncertain [26]. Hence, it is necessary to establish the lower tolerance limit of the fatigue life at a given failure probability p. Here, we employed a pairwise bootstrap [11] for the lower tolerance limit of the *P-S-N* curve obtained from the QR model. Specifically, a bootstrap data set,$(x_1^*,y_1^*),(x_2^*,y_2^*),\cdots,(x_n^*,y_n^*)$, is first generated by random sampling with replacement from the test data(x,y). Then the bootstrap regression coefficient vector.

$$\widehat{\boldsymbol{\beta}}^*(p) = \underset{\beta(p)}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_p(y_i^* - \mathrm{x}_i^* \beta(p))$$

are calculated. Furthermore, the p quantile of fatigue life $\widehat{Q}(p)$ can estimated. Repeating above process r times we can obtain r$\widehat{Q}(p)$, and its lower α percentile is approximated as the lower tolerance limit. Fig. 7 shows the 1 % failure probability *S-N* curve with 95 % confidence tolerance bound in log–log coordinate. It can be seen that the gap between the 95 % confidence tolerance bound and the 50 % tolerance bound is very small, which indicates sample size at each stress level is enough for the QR model which is quite robust.

5.2. S420MC steel

The QR model was also verified on an S420MC steel dataset, and the detailed experimental process was described in [45]. A total of 65 specimens broken before 2 million load cycles are applied to estimate the S420MC *S-N* curves and its scatter. The experimental data and *S-N* curves with different failure probabilities are described in log–log coordinate (Fig. 8). Fig. 8a and b give the predicted probability *S-N* curves of the QR model and the ISO method, respectively. It can be seen that for both models, around 50 %, 95 %, and 99 % of the test fatigue failure data are above median *S-N* curve, 5 % failure probability *S-N* curve, and 1 % failure probability *S-N* curve, respectively. More specifically, the ISO model is more conservative than the QR model when the failure probability is 1 %. In addition, we can see all the *P-S-N* curves are parallel. According the Bartlett test, the chi-square value is 9.862, the freedom degree is 11, and the p-value is 0.5428. This means the null hypothesis cannot be rejected that the variances across all the stress levels are homogeneous. In the Bartlett test, at least two observations are required for each group, so a total of 60 specimens from 12 stress levels are tested.

The reliability of the S420MC steel at some stress levels are presented in Fig. 9. We can see that both models are more conservative than the K-M estimator when the reliabilities are high. For 232 and 250 stress levels, when reliabilities are between 50 % and 80 %, the service cycles estimated by the QR model are similar to the K-M estimators while the estimators of the ISO model are more radical than the K-M estimators.

Fig. 10 shows the 1 % failure probability *S-N* curve with a 95 % confidence tolerance bound in the log–log coordinate. As we can see the gaps between the 95 % confidence tolerance bound and the 50 % tolerance bound initially decreases when the stress level decreases until down to 200 MPa, and then the gap continues to increase as the decreasing of the stress level. This means the uncertainty of the quantile regression at higher stress levels and the low-stress level outside of the range of the experimental data is larger than that at the stress level around 200 MPa, which may cause by less arranged tests. For more specific, a total of 25 specimens are tested between 204 MPa and 229 MPa, and 63 specimens are tested between 232 MPa and 295 MPa.

6. Conclusions and discussions

In this work, we present a novel framework for the estimation of probabilistic *S-N* curves in fatigue of materials based on the extended quantile regression (QR) model. We firstly extended the traditional QR model by imposing some linear inequality constraints to ensure some physical sounds in applications of fatigue problems of materials, e.g. non-crossing and monotonic of the estimated multiple probabilistic *S-N* curves. And the applicability of the extended QR model was then verified with some different conditions: simulation data with different underlying distributions, homoscedasticity and heteroscedasticity, and experimental datasets with different engineering material types. In compared with the currently adopted estimation model of probabilistic *S-N* curves, the QR model can be built directly on the experimental data without any distribution assumption for the fatigue life. The massive numerical comparisons demonstrate that the QR model is exempt from the problem of incorrectly specifying the potential fatigue life distribution, and is robust to the non-constant scale problem.

To approach the extended QR model using for the estimation of probabilistic *S-N* curves in fatigue of materials, there are several issues merit further exploration. The first one is regarding the application conditions of the proposed model to the fatigue problems of real engineering materials. To address this issue, we investigate the performance of the proposed model with different conditions: three distributions (Lognormal, Weibull, and log-logistic distributions), two variance scenarios (homoscedasticity and heteroscedasticity), and two different fatigue datasets of metallic materials (the aluminum alloy 2524-T4 and the S420MC steel). The proposed model shows great applicability and robustness under those conditions. The applicability of the proposed model can be further verified by fatigue datasets under more different conditions, such as, fatigue operations, engineering materials, etc. Secondly, the extensions in the model may further be explored. In this paper, we employed the simplest Basquin model to build the relationships between fatigue life and stress level, and noting that the model has better performance effect on medium and large database. Some small sample methods can be further explored in the framework of quantile regression since the fatigue test usually being time-consuming and costly.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors are very grateful to the editors and reviewers for their valuable comments and suggestions which helped us to improve this paper. Besides, we would like to acknowledge support from the Science Challenge Project (No. TZ2018002), the National Natural Science Foundation of China (No. 12002343), the R&D Program of Beijing Municipal Education Commission (KM202011232018), and the Key Research and Cultivation Project of Scientific Research on Campus of Beijing Information Science and Technology University in 2021 (2021YJPY236).

Appendix A

See Figs. A1 and A2, A3 and A4.

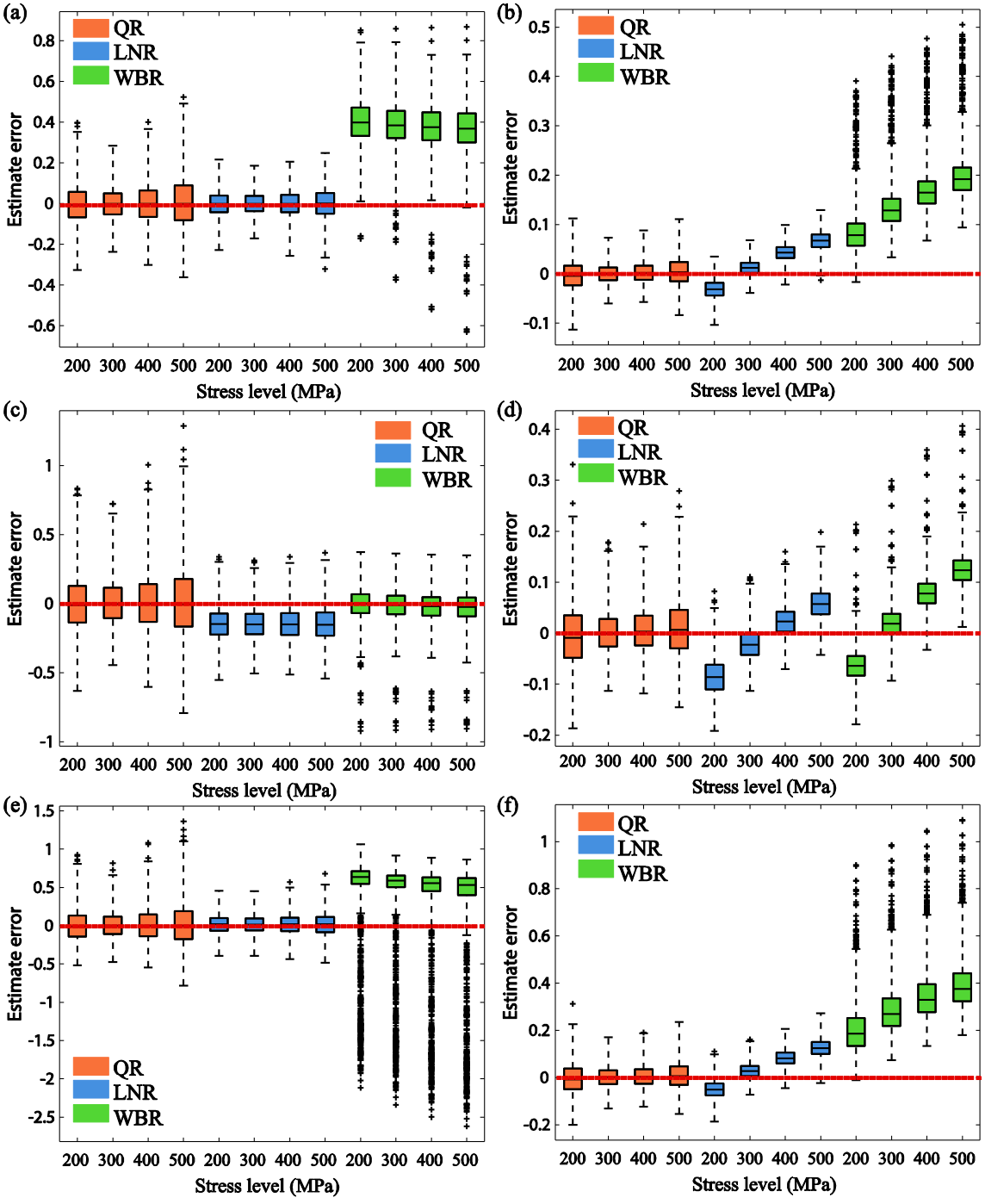

Fig. A1. Boxplot of the prediction errors at 5% failure probability under four applied stress levels, and 50 samples are drawn at each stress level. The left panel shows the prediction errors in the constant scale scenario, and the right panel are the prediction errors in the non-constant scale scenario.

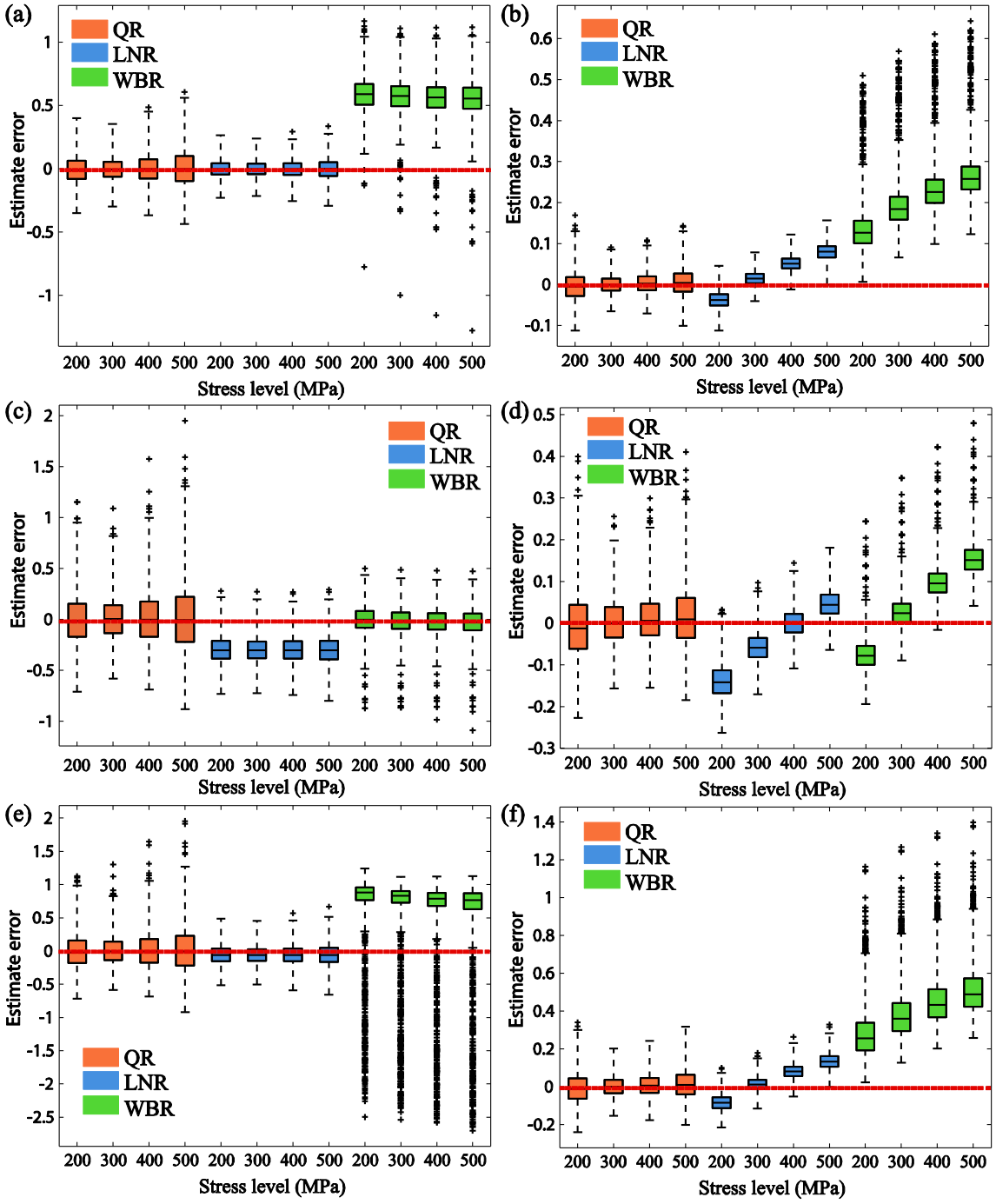

Fig. A2. Boxplot of the prediction errors at 2.5% failure probability under four applied stress levels, and 50 samples are drawn at each stress level. The left panel shows the prediction errors in the constant scale scenario, and the right panel are the prediction errors in the non-constant scale scenario.

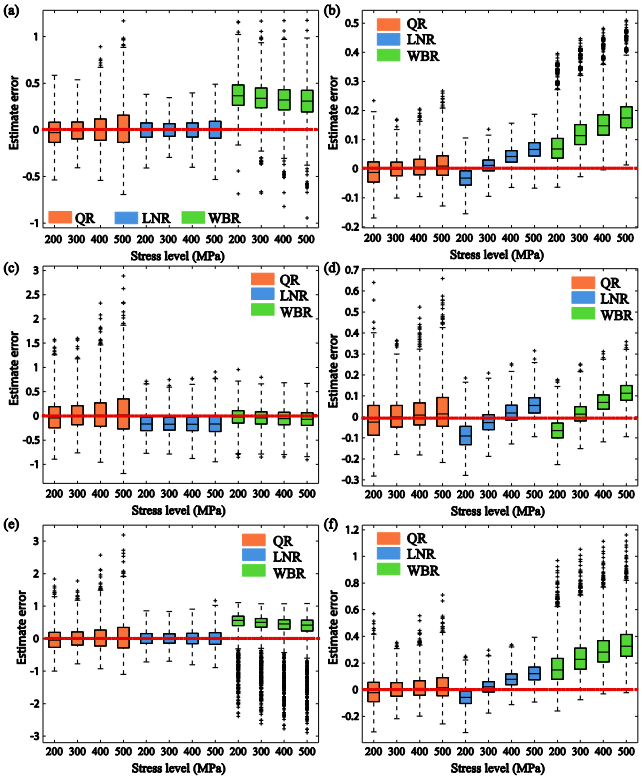

Fig. A3. Boxplot of the prediction errors at 5% failure probability under four applied stress levels, and 15 samples are drawn at each stress level. The left panel shows the prediction errors in the constant scale scenario, and the right panel are the prediction errors in the non-constant scale scenario.

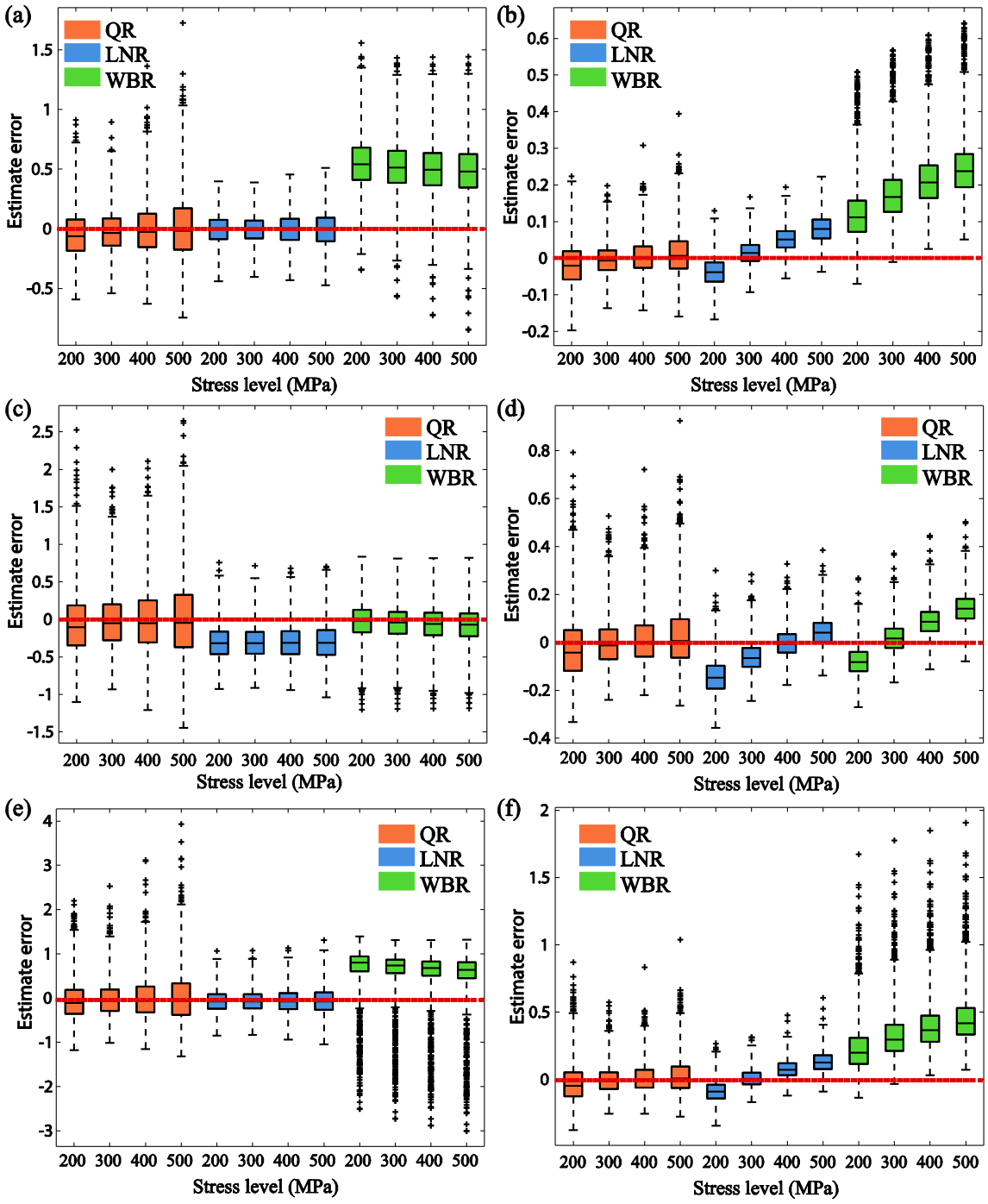

Fig. A4. Boxplot of the prediction errors at 2.5% failure probability under four applied stress levels, and 15 samples are drawn at each stress level. The left panel shows the prediction errors in the constant scale scenario, and the right panel are the prediction errors in the non-constant scale scenario.

References

[1] Barrodale I, Roberts FDK. Solution of an overdetermined system of equations in the l_1 norm. Commun ACM 1974;17(6):319–20. https://doi.org/10.1145/355616.361024.

[2] Bartlett MS. Properties of sufficiency and statistical tests. Proc Roy Soc London. Series A-Math Phys Sci 1937; 160(901): 268-282. https://doi.org/10.1098/rspa.1937.0109.

[3] Basquin OH. The exponential law of endurance tests. Am Soc Testing Mater 1910; 10:625–30.

[4] Castillo E, López-Aenlle M, Ramos A, et al. Specimen length effect on parameter estimation in modelling fatigue strength by Weibull distribution. Int J Fatigue 2006;28(9):1047–58. https://doi.org/10.1016/j.ijfatigue.2005.11.006.

[5] Chen J, Diao B, He J, et al. Equivalent surface defect model for fatigue life prediction of steel reinforcing bars with pitting corrosion. Int J Fatigue 2018;110: 153–61. https://doi.org/10.1016/j.ijfatigue.2018.01.019.

[6] Chen J, Liu S, Zhang W, et al. Uncertainty quantification of fatigue S-N curves with sparse data using hierarchical Bayesian data augmentation. Int J Fatigue 2020;134: 105511. https://doi.org/10.1016/j.ijfatigue.2020.105511.

[7] Chen J, Liu Y. Probabilistic physics-guided machine learning for fatigue data analysis. Expert Syst Appl 2021;168:114316. https://doi.org/10.1016/j.eswa.2020.114316.

[8] Collins JA. Failure of materials in mechanical design: analysis, prediction, prevention. John Wiley & Sons; 1993.

[9] Dolan T J, Brown H F. Effect of prior repeated stressing on fatigue life of 75S T aluminum. Department of Theoretical and Applied Mechanics. College of Engineering. University of Illinois at Urbana-Champaign, 1952.

[10] DuQuesnay DL, Underhill PR. Fatigue life scatter in 7xxx series aluminum alloys. Int J Fatigue 2010;32(2):398–402. https://doi.org/10.1016/j.ijfatigue.2009.07.016.

[11] Efron B, Tibshirani R J. An introduction to bootstrap, 1994, CRC press.

[12] Fouchereau R, Celeux G, Pamphile P. Probabilistic modeling of S-N curves. Int J Fatigue 2014;68:217–23. https://doi.org/10.1016/j.ijfatigue.2014.04.015.

[13] Guan X, He J. Life time extension of turbine rotating components under risk constraints: A state-of-the-art review and case study. Int J Fatigue 2019;129: 104799. https://doi.org/10.1016/j.ijfatigue.2018.08.003.

[14] Hoole J, Sartor P, Booker J, et al. Systematic statistical characterisation of stress-life datasets using 3-Parameter distributions. Int J Fatigue 2019;129:105216. https://doi.org/10.1016/j.ijfatigue.2019.105216.

[15] ISO 12107. Metallic materials-Fatigue testing-Statistical planning and analysis of data-ISO 12107.Test, 2003. p. 2.

[16] Júnior RCSF, Belísio AS. Probabilistic S-N curves using exponential and power laws equations. Compos B Eng 2014;56:582–90. https://doi.org/10.1016/j.compositesb.2013.08.036.

[17] Kracík J, Strnadel B. A statistical model for lifespan prediction of large steel structures. Eng Struct 2018;176:20–7. https://doi.org/10.1016/j.engstruct.2018.08.065.

[18] Koenker R, Bassett Jr G. Regression quantiles. Econometrica 1978:33–50. https://doi.org/10.2307/1913643.

[19] Koenker R. Quantile regression. Cambridge, U.K.: Cambridge University; 2005.

[20] Liu XW, Lu DG, Hoogenboom PCJ. Hierarchical Bayesian fatigue data analysis. Int J Fatigue 2017;100:418–28. https://doi.org/10.1016/j.ijfatigue.2017.03.043.

[21] Liu Y, Mahadevan S. Probabilistic fatigue life prediction using an equivalent initial flaw size distribution. Int J Fatigue 2009;31(3):476–87. https://doi.org/10.1016/j.ijfatigue.2008.06.005.

[22] Mohabeddine A, Correia J, Montenegro PA, et al. Probabilistic SN curves for CFRP retrofitted steel details. Int J Fatigue 2021;148:106205. https://doi.org/10.1016/j.ijfatigue.2021.106205.

[23] Muggeo VMR, Sciandra M, Tomasello A, et al. Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. Environ Ecol Stat 2013;20(4):519–31. https://doi.org/10.1007/s10651-012-0232-1.

[24] Pascual FG, Meeker WQ. Estimating fatigue curves with the random fatigue-limit model. Technometrics 1999;41(4):277–89. https://doi.org/10.1080/00401706.1999.10485925.

[25] Pascual FG, Montepiedra G. Lognormal and Weibull accelerated life test plans under distribution misspecification. IEEE Trans Reliab 2005;54(1):43–52. https://doi.org/10.1109/TR.2004.837316.

[26] Ronold KO, Lotsberg I. On the estimation of characteristic S-N curves with confidence. Mar struct 2012;27(1):29–44. https://doi.org/10.1016/j.marstruc.2012.03.002.

[27] Shen C. The statistical analysis of fatigue data. The University of Arizona; 1994.

[28] Shimokawa T, Hamaguchi Y. Relationship between fatigue life distribution, notch configuration, and SN curve of a 2024–T4 aluminum alloy. J Eng Mater Technol 1985;107(3):214–20. https://doi.org/10.1115/1.3225804.

[29] Sun C, Song Q, Hu Y, Wei Y. Effects of intermittent loading on fatigue life of a high strength steel in very high cycle fatigue regime. Int J Fatigue 2018;117:9–12. https://doi.org/10.1016/j.ijfatigue.2018.07.033.

[30] Sun C, Song Q, Zhou L, Liu J, Wang Y, Wu X, et al. The formation of discontinuous gradient regimes during crack initiation in high strength steels under very high cycle fatigue. Int J Fatigue 2019;124:483–92. https://doi.org/10.1016/j.ijfatigue.2019.03.026.

[31] Tanaka S, Akita S. On the miner's damage hypothesis in notched specimens with emphasis on scatter of fatigue life. Eng Fract Mech 1975;7(3):473–80. https://doi.org/10.1016/0013-7944(75)90048-X.

[32] Tan X. P-S–N curve fitting method based on sample aggregation principle. J Fail Anal Prev 2019;19(1):270–8. https://doi.org/10.1007/s11668-019-00586-1.

[33] Vassilopoulos AP, Georgopoulos EF, Keller T. Comparison of genetic programming with conventional methods for fatigue life modeling of FRP composite materials. Int J Fatigue 2008;30(9):1634–45. https://doi.org/10.1016/j.ijfatigue.2007.11.007.

[34] Wang C, Wen J, et al. Anisotropic expansion and size-dependent fracture of silicon nanotubes during lithiation. J Mater Chem A 2019;7(25):15113–22. https://doi.org/10.1039/C9TA00519F.

[35] Weibull W A. Statistical theory of the strength of materials. Proceedings of Royal Swedish Institute Engineering Researching, 1939, 1-50.

[36] Wen J, Wei Y, Cheng YT. Stress evolution in elastic-plastic electrodes during electrochemical processes: A numerical method and its applications. J Mech Phys Solids 2018;116:403–15. https://doi.org/10.1016/j.jmps.2018.04.006.

[37] Wen J, Zou Q, Zhang Z, et al. The scaling of charging rate and cycle number of commercial batteries. Acta Mech Sin 2022;38(5):1–10. https://doi.org/10.1007/s10409-022-22108-x.

[38] Wu Y, Liu Y. Stepwise multiple quantile regression estimation using non-crossing constraints. Statist Interface 2009;2(3):299–310. https://doi.org/10.4310/SII.2009.v2.n3.a4.

[39] Xie L, Liu J, Wu N, et al. Backwards statistical inference method for P-S–N curve fitting with small-sample experiment data. Int J Fatigue 2014;63:62–7. https://doi.org/10.1016/j.ijfatigue.2014.01.006.

[40] Yuan Z, Chen X, Ma L, Li Q, Sun S, Wei Y. A segmented load spectrum model for high-speed trains and its inflection stress as an indicator for line quality. Int J Fatigue 2021;148:106221. https://doi.org/10.1016/j.ijfatigue.2021.106221.

[41] Zhao YX, Gao Q, Wang JN. An approach for determining an appropriate assumed distribution of fatigue life under limited data. Reliab Eng Syst Saf 2000;67(1):1–7. https://doi.org/10.1016/S0951-8320(99)00036-8.

[42] Zu T, Kang R, Wen M, et al. α-SN curve: a novel SN curve modeling method under small-sample test data using uncertainty theory. Int J Fatigue 2020;139:105725. https://doi.org/10.1016/j.ijfatigue.2020.105725.

[43] Nabizadeh A, Tabatabai H. Development of nonlinear probabilistic SN curves using survival analysis techniques with application to steel bridges. Int J Fatigue 2020; 141:105892. https://doi.org/10.1016/j.ijfatigue.2020.105892.

[44] Pan X, Su H, Sun C, Hong Y. The behavior of crack initiation and early growth in high-cycle and very-high-cycle fatigue regimes for a titanium alloy. Int J Fatigue 2018;115:67–78. https://doi.org/10.1016/j.ijfatigue.2018.03.021.

[45] Klemenc J, Fajdiga M. Estimating S–N curves and their scatter using a differential ant-stigmergy algorithm. Int J Fatigue 2012;43:90–7. https://doi.org/10.1016/j.ijfatigue.2012.02.015.