

Empacher, Christina; Kamps, Udo; Schmiedt, Anja Bettina

**Article — Published Version**

## Prediction intervals for future Pareto record claims

European Actuarial Journal

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Empacher, Christina; Kamps, Udo; Schmiedt, Anja Bettina (2024) : Prediction intervals for future Pareto record claims, European Actuarial Journal, ISSN 2190-9741, Springer, Berlin, Heidelberg, Vol. 15, Iss. 1, pp. 163-197, <https://doi.org/10.1007/s13385-024-00397-1>

This Version is available at:

<https://hdl.handle.net/10419/323663>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# Prediction intervals for future Pareto record claims

Christina Empacher<sup>1</sup> · Udo Kamps<sup>1</sup> · Anja Bettina Schmiedt<sup>2</sup> 

Received: 17 November 2023 / Revised: 29 July 2024 / Accepted: 1 August 2024 /

Published online: 7 October 2024

© The Author(s) 2024

## Abstract

Stochastic models and methods for quantifying extreme events are of interest in numerous disciplines. Within this paper, statistical prediction of extreme claims or losses in insurance industry is considered based on upper record values, which describe successively largest observations in a sequence of data over time. The problem of predicting a future record value (here in particular, a future record claim) based on a sequence of previously observed record values (here, past record claims) is addressed by means of prediction intervals. For an underlying Pareto distribution, respective exact and approximate intervals from the literature are summarized and modified and new ones are developed. In a simulation study, these prediction intervals are evaluated and compared regarding coverage frequency and length. The impact of the number of observed record values as well as the choice of the Pareto distribution is discussed. In the case of a small number of record values, the use of  $k$ -th record values is considered as an option for statistical analyses to predict, e.g., second largest record claims. Selected prediction methods are applied to several real data sets, which turn out to perform well and to be able to capture the magnitude of future record claims, even for fairly small numbers of record observations. For comparison, generalized Pareto distributions are fitted to real data sets and a corresponding point predictor as well as a respective upper prediction interval for the next record to appear are derived and evaluated.

**Keywords** Record values · Interval prediction · Pareto distribution · Generalized Pareto distribution · Real insurance data sets

---

Christina Empacher, Udo Kamps and Anja Bettina Schmiedt have contributed equally to this work.

---

✉ Anja Bettina Schmiedt  
anja.schmiedt@th-rosenheim.de; anja.bettina.schmiedt@gmail.com

Christina Empacher  
empacher@isw.rwth-aachen.de

Udo Kamps  
udo.kamps@rwth-aachen.de

<sup>1</sup> Institute of Statistics, RWTH Aachen University, Aachen 52056, Germany

<sup>2</sup> Department of Mathematics, Rosenheim Technical University of Applied Sciences, Rosenheim 83024, Germany

**JEL Classification** C130 · C150 · C460 · C530 · C800 · G220

## 1 Introduction

### 1.1 Motivation

In recent decades, a strong need has manifested itself in statistical models for quantifying rare and extreme events. Such models are of interest in numerous disciplines, for example in environmental sciences, engineering, finance and insurance. In the insurance industry, statistical predictions of extreme claims resp. losses are crucial for pricing and for adequate quantitative risk models. The latter should provide a sound assessment of the risk-bearing capacity of an insurance company, which is economically relevant as well as required by supervisory law. As a consequence, the probabilistic behavior of extreme claim amounts plays a decisive role in the actuarial departments of insurance companies, e.g. in the development of risk mitigation strategies, in the pricing of reinsurance contracts and the evaluation of inflation effects.

The subject of the present paper is statistical modeling and prediction of extreme claims resp. losses in non-life insurance. The analyzed random variables are thus claim or loss sizes, where the terms are used as synonyms in most applications. Then, future largest claim sizes are to be predicted on the basis of successively largest observed claim sizes in the past. More precisely, the statistical method that we introduce is based on sequences of upper record values. In contrast to classical extreme value methods, the development and application of statistical methods based on records do not seem to have been considered in actuarial risk analysis.

The study of record statistics dates back to Chandler [16]. Record values constitute a canonical model for the temporal development of extreme values, such as extreme claim sizes in insurance. Castaño-Martínez et al. [15] consider record values in a sequence of independent and identically distributed claims and introduce a respective premium principle. The topic of prediction of record values as considered in the present paper may also be applied in environmental studies, such as prediction of rainfall extremes, highest water levels, or air record temperatures [cf. 27] as well as in sport analytics [cf. 19]. Meanwhile, there is a broad literature on record values, where basic probabilistic and statistical results can be found in the monographs by Arnold et al. [4] and Nevzorov [25].

We provide an introduction to the model of record statistics in Sect. 1.2. Following the notation therein, we are concerned with the problem of predicting the value of a future record (future record claim)  $R_s$  based on the first  $r$  observed record values (past record claims)  $R_1, \dots, R_r$ , where  $s > r$ .

From a practical point of view, we are thus interested in smallest possible prediction intervals that contain future record claim sizes  $R_s$ ,  $s > r$ , with a given high probability. That is, we will focus on interval prediction rather than point prediction for future record claims, since intervals are more meaningful in practice.

In the record model, record values are understood as a natural model for the sequence of successive extremes in a sequence of independent and identically distributed (iid) non-negative random variables. In our applications, the underlying sequence of respec-

tive observations consists of successively observed non-life insurance claims resp. losses over time along with the appearing record values, in both, simulated and several real data sets.

Techniques from extreme value analysis are well established in actuarial science (cf., e.g., Beirlant et al. [9], Embrechts et al. [18]). In situations such as pricing a high-excess layer in reinsurance, the modeling of tails of loss distributions is usually based on extreme value theory. Applications of extreme value statistics in modeling and predicting large claims in insurance can be found, e.g., in Beirlant and Teugels [8], Embrechts et al. [18, Chapter 6] and McNeil [23], see also the monograph by Beirlant et al. [9, Section 1.3, Section 6.2]. By fitting either an extreme value distribution to maximum losses or a generalized Pareto distribution to losses exceeding a certain threshold, forecasting methods of classical extreme value theory can be seen as long-term prediction by specifying high quantiles, such as the 99.5%-quantile. The objective of the proposed approach in the present paper, on the other hand, is the statistical point and interval prediction of future record claims based on the information that is given by successive observed largest claims in the past. In this sense, the study and prediction of record values is understood as a supplementing and decision-supporting tool in risk analysis and internal risk assessment.

Regarding the record model, the statistics used for prediction of future records depend highly on the assumed underlying distribution. Moreover, we focus on providing explicit methods for interval prediction of future record claims. In the present paper, we consider Pareto as underlying distribution, which in actuarial science and non-life applications is frequently considered for modeling heavy-tail situations or data above a threshold; see, e.g., Resnick [30], Bladt et al. [12], Albrecher et al. [3]. Pareto and generalized Pareto distributions are, with an appropriate choice of distribution parameters, considered as heavy-tailed distributions and are thus suitable for modeling extreme events. Their use for modeling claims severity and measuring tail risks has also been discussed by, e.g., Brazauskas and Kleefeld [14] with application to Norwegian fire claims and by, e.g., Brazauskas and Kleefeld [13] and Raschke [29] with application to Danish fire losses. In the literature, the application of (generalized) Pareto distributions to high extremes of Danish fire losses is well known by the paper of McNeil [23] and the corresponding discussion by Resnick [30]. In the real data analyses in Sect. 4, we will consider the Danish fire losses as well as other real data sets to validate and discuss the developed record prediction methods.

In order to fit a Pareto distribution to highest claims, the choice of the threshold is important. As discussed by Resnick [30, p. 140], graphical tools such like the Hill plot might be 'more guesswork than science'. Authors try to handle the threshold by suggesting more robust procedures to estimate distribution parameters including the threshold, see, e.g., Brazauskas and Kleefeld [13]. In the data analyses in the present paper, we choose a different approach and assume that severe claim sizes above a certain threshold are given and that Pareto distributions with unknown shape parameter are appropriate. The Danish fire losses, for example, have been anyway observed above a fairly large threshold.

Data from two recent real data examples on flood losses and man-made catastrophe losses above some threshold turn out to be very well represented by Pareto distributions. From extreme value theory it is well known that generalized Pareto distributions

(GPDs) appear as limit distributions of scaled quantities exceeding high thresholds; GPDs may have a better fit to the considered data sets. We examine respective fittings and resulting predictions in the applications.

## 1.2 Record statistics

In a sequence of observations of insurance claims over time, say, upper record values are nothing but successively largest observations within this sequence. The mathematical model is as follows. Upper record values in a sequence of independent and identically distributed random variables  $X_1, X_2, \dots$  with absolutely continuous distribution function  $F$  and probability density function (pdf)  $f$  are defined via record times

$$L(1) = 1, L(n+1) = \min \{j > L(n) | X_j > X_{L(n)}\}, \quad n \in \mathbb{N}, \quad (1)$$

i.e., random numbers at which record values appear, as

$$R_n = X_{L(n)}, \quad n \in \mathbb{N}$$

[see 4, 25]. Hence, record values constitute an increasing random subsequence of the original random sequence  $X_1, X_2, \dots$ , the latter of which, in our application, models successive non-life insurance claims in predefined time periods. In Castaño-Martínez et al. [15, Section 3.3], the expected value of  $R_n$  is introduced and examined as a distortion premium principle, where the index  $n$  may be interpreted as degree of risk aversion of the insurer.

A single record with an extraordinary large value in the very beginning of a sequence of data over time may lead to only a few observed records afterwards. Here,  $k$ -th record values are a suitable model in order to describe the behavior of records over time. Moreover, when dealing with a fixed number of record observations in a sequence of data, there will be more observations in the respective series of  $k$ -th records,  $k \geq 2$ , than in the series of common record values with  $k = 1$ , which may be reasonable regarding precision of estimators and predictors. Therefore, if not the largest values, but successively  $k$ -th largest record values  $R_n^{(k)}$ ,  $n \in \mathbb{N}$ , for some fixed  $k \in \mathbb{N}$  are of particular interest, then these are defined via  $k$ -largest order statistics, denoted by  $X_{j,j+k-1}$  in the following, by means of

$$R_n^{(k)} = X_{L^{(k)}(n), L^{(k)}(n)+k-1}, \quad n \in \mathbb{N}, \quad (2)$$

with record times

$$L^{(k)}(1) = 1, \\ L^{(k)}(n+1) = \min \left\{ j > L^{(k)}(n) | X_{j,j+k-1} > X_{L^{(k)}(n), L^{(k)}(n)+k-1} \right\}, \quad n \in \mathbb{N}$$

[cf. 4, 17, 21]. In the context of insurance mathematics, Castaño-Martínez et al. [15, Section 3.4] introduce a distortion premium principle based on  $k$ -th largest record

values in order to gain flexibility in modeling risk aversion. In general, for  $k = 1$ , common record times  $L(n) = L^{(1)}(n)$  and record values  $R_n = R_n^{(1)}$  are included in the model.

**Remark 1** It is well known that  $k$ -th record values with an underlying distribution function  $F$  are identically distributed as common record values based on the distribution function  $G = 1 - (1 - F)^k$  [cf. 24]. Respective distributional properties of  $k$ -th record values may thus be obtained from common record values [cf. 17].

In the literature [see, e.g., 4], we find the joint density function of the first  $r$  record values

$$f^{R_1, \dots, R_r}(x_1, \dots, x_r) = \left( \prod_{i=1}^{r-1} \frac{f(x_i)}{1 - F(x_i)} \right) f(x_r), \quad x_1 \leq \dots \leq x_r,$$

the marginal density function

$$f^{R_r}(x) = \frac{1}{(r-1)!} (-\log(1 - F(x)))^{r-1} f(x), \quad x \in \mathbb{R}, \quad r \in \mathbb{N}, \quad (3)$$

and the marginal distribution function

$$F^{R_r}(x) = 1 - (1 - F(x)) \sum_{j=0}^{r-1} \frac{1}{j!} \left( -\log(1 - F(x)) \right)^j, \quad x \in \mathbb{R}, \quad r \in \mathbb{N},$$

of the  $r$ -th record value.

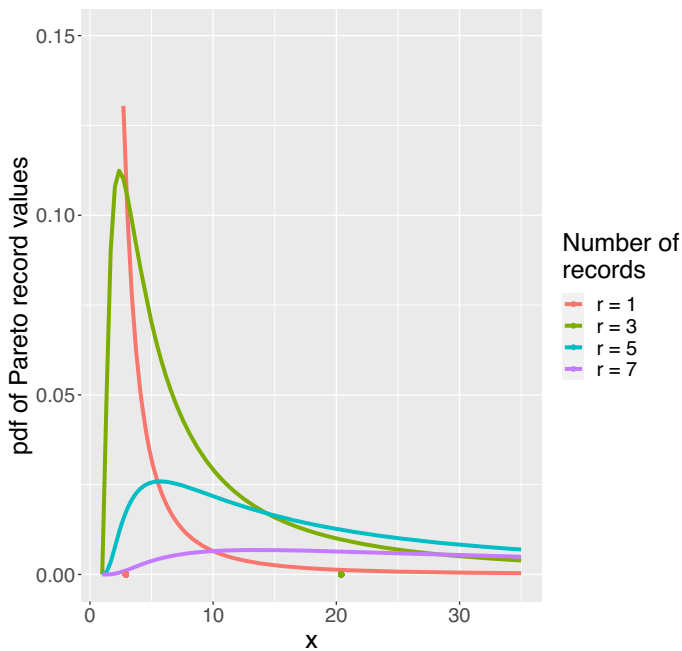
Moreover, upper record values form a Markov chain with transition probabilities

$$P(R_r > t | R_{r-1} = s) = \frac{1 - F(t)}{1 - F(s)}, \quad s \leq t, \quad F(s) < 1, \quad (4)$$

for  $r \geq 2$ .

A visualization of the marginal densities of record values as in (3) can be found in Fig. 1 based on a Pareto distribution with density function  $f$  given by  $f(x) = 1.3 x^{-2.3}$ ,  $x > 1$ . First, it is worth noticing that the pdf of the first record  $R_1$  in Fig. 1 coincides with the underlying pdf  $f$ . In particular, it can be seen that the mass of the distribution is shifted to the right as  $r$  increases. Note that all figures in the present paper were created using the statistical software R.

In order to further demonstrate the rapid growth of quantiles of the distributions of subsequent record values, medians, upper quartiles and 90%-quantiles of the distributions of record values are shown in Table 1, again based on a Pareto distribution with density function  $f(x) = 1.3x^{-2.3}$ ,  $x > 1$ .



**Fig. 1** Density functions of record values  $R_r$  with  $r \in \{1, 3, 5, 7\}$  based on a Pareto distribution with pdf  $f(x) = 1.3x^{-2.3}$ ,  $x > 1$ . Upper quartiles for  $r = 1$  and  $r = 3$  are indicated by dots

**Table 1** Quantiles of the distributions of record values  $R_r$  as in Fig. 1 for different values of  $r$

$r$	median	upper quartile	90%-quantile
1	1.7	2.9	5.9
2	3.6	7.9	19.9
3	7.8	20.4	60.0
4	16.9	50.9	170.6
5	36.3	124.8	468.2
6	78.4	301.8	1254.3
7	169.1	723.0	3299.7

### 1.3 Outline

In Sect. 2, point and interval prediction of future record values based on  $r$  previous record values  $R_1, \dots, R_r$  for some  $r \in \mathbb{N}$  are shown. Some exact and approximate prediction intervals are known from the literature, the latter of which are modified here with respect to estimating an unknown distribution parameter based on all available data above some threshold and not on record values, only. This approach is reasonable in the insurance application. Two more prediction intervals are developed based on a point predictor introduced in Volovskiy and Kamps [32]. Results from an extensive simulation study are presented in Sect. 3, where one-sided prediction intervals are

compared by means of coverage frequencies and mean lengths. As a consequence, selected methods are recommended for practical use. These are applied in Sect. 4 to real insurance data sets and the outcomes are discussed. A point predictor of the next record value and a prediction interval based on a GPD are introduced in Sect. 5. For comparison, such GPDs are fitted to real data sets and the respective prediction results are shown.

## 2 Point and interval prediction for future record values

In this section, we briefly discuss point prediction and focus on prediction intervals for future record values based on observed record values  $R_1, \dots, R_r$  for some  $r \in \mathbb{N}$ . For previous works on this topic, we refer to, e.g., Awad and Raqab [6], Raqab et al. [28], Asgharzadeh et al. [5] and Empacher et al. [19]. After stating point predictors and prediction intervals known from the literature and modifying these with respect to an application on successive largest claims in sequences of claims over time in non-life insurance, we mainly supplement these methods by two new prediction intervals for the next record value  $R_{r+1}$ . We are basically interested in upper prediction intervals with  $R_r$  as lower bound, which are compared and applied in Sect. 3. The prediction methods are based on record values from an underlying Pareto distribution  $Par(\lambda, \beta)$ , with scale parameter  $\lambda > 0$  and shape parameter  $\beta > 0$  with distribution function

$$F(x) = 1 - \left(\frac{\lambda}{x}\right)^\beta, \quad x \in [\lambda, \infty), \quad (5)$$

and density function

$$f(x) = \frac{\beta \lambda^\beta}{x^{\beta+1}}, \quad x \in [\lambda, \infty).$$

In Section 4, we apply our methods to real data of insurance (record) claims by setting a threshold  $\lambda > 0$  in order to approximately meet the underlying iid assumption and to fit a Pareto distribution. Thus, in what follows, the scale parameter  $\lambda$  is assumed to be known. In Subsection 2.1, we first deal with point prediction of future record values.

**Remark 2** For a Pareto distribution as in (5), it suffices to develop statistical methods for common records, only, since  $1 - (1 - F)^k$  is the distribution function of  $Par(\lambda, \beta k)$ . Thus, considering  $k$ -th record values based on  $Par(\lambda, \beta)$  amounts to applying common record values from  $Par(\lambda, \beta k)$ .

For point prediction of  $k$ -th record values and Bayesian methods, we refer to Ahmadi et al. [2].



## 2.1 Point prediction of future Pareto record values

Likelihood based prediction is frequently studied in the literature. Let  $\theta$  denote the unknown parameter or parameter vector of the underlying distribution of record values with absolutely continuous distribution function  $F_\theta$  and density function  $f_\theta$ . A maximum likelihood predictor (MLP) of  $R_s$  based on  $R_1, \dots, R_r$  is derived by maximizing the joint density function

$$f_\theta^{R_1, \dots, R_r, R_s}(r_1, \dots, r_r, r_s)$$

of  $R_1, \dots, R_r$  and  $R_s$ ,  $s > r$ , with respect to  $r_s$  and  $\theta$ ; an optimal solution of  $r_s$  defines the MLP and the optimal solution  $\hat{\theta}$  of  $\theta$  is called predictive maximum likelihood estimator of  $\theta$ .

In order to obtain a maximum observed likelihood predictor (MOLP) of  $R_s$  given  $R_1, \dots, R_r$ , the conditional density function

$$f_\theta^{R_1, \dots, R_r | R_s}(r_1, \dots, r_r | r_s)$$

of  $R_1, \dots, R_r$  given  $R_s$  is maximized with respect to  $r_s$  and  $\theta$ ; then, an optimal solution of  $r_s$  defines the MOLP (cf. Bayarri and DeGroot [7]).

Based on a Pareto distribution as in (5), the MOLP is uniquely determined by

$$\hat{R}_{sMOLP} = R_1 \left( \frac{R_r}{R_1} \right)^{\frac{s-1}{r}} \quad \text{and} \quad {}_\lambda \hat{R}_{sMOLP} = \lambda \left( \frac{R_r}{\lambda} \right)^{\frac{s-1}{r}}$$

for an unknown and a known scale parameter  $\lambda > 0$ , respectively [see 7, 31, 33].

The MLP of  $R_s$  based on the record values  $R_1, \dots, R_r$ ,  $s > r$ , turns out to be more involved and is uniquely given by

$$\hat{R}_{sMLP} = R_r \exp \left\{ \frac{s-r-1}{\hat{\beta}+1} \right\}$$

with  $\hat{\beta} = \frac{1}{2} \left( \frac{r+1}{\tilde{R}} - 1 + \left( \left( \frac{r+1}{\tilde{R}} - 1 \right)^2 + \frac{4s}{\tilde{R}} \right)^{1/2} \right)$  and  $\tilde{R} = \ln R_r - \ln R_1$  in case  $\lambda$  is unknown [see 28, 33]. If  $\lambda$  is known, then  $R_1$  has to be replaced by  $\lambda$  to obtain the MLP. In Volovskiy and Kamps [33], the predictors  $\hat{R}_{sMLP}$  and  $\hat{R}_{sMOLP}$  are compared by means of Pitman measure of closeness for  $s > r + 1$ , where the maximum observed likelihood predictor  $\hat{R}_{sMOLP}$  outperforms the maximum likelihood predictor  $\hat{R}_{sMLP}$ .

However, both maximum-likelihood based predictors share the same drawback in the case  $s = r + 1$ ; namely, both predictors coincide with the last observed record value  $R_r$  and hence lead to trivial predictors.

In Volovskiy and Kamps [32], the maximum-product-of-spacings predictor (MPSP) of  $R_s$ ,  $s > r$ , is introduced as an alternative. It is given by

$$\hat{R}_{sMPSP} = R_1 \left( \frac{R_r}{R_1} \right)^{\frac{s-1}{r-1}}, \quad \text{if } \lambda \text{ is unknown, and}$$

$${}_{\lambda}\hat{R}_{sMPS} = \lambda \left( \frac{R_r}{\lambda} \right)^{s/r}, \text{ if } \lambda \text{ is known,} \quad (6)$$

and is therefore useful to predict the next future record value  $R_{r+1}$  (i.e.,  $s = r + 1$ ). Obviously, the derivation of the predictors is independent of the shape parameter  $\beta$ .

For Bayesian prediction methods based on upper Pareto record values, we refer to Ahmadi and Doostparast [1], Madi and Raqab [22] and Raqab et al. [28].

**Remark 3** The expected values of the future record value  $R_s, s > r$ , and of its predictor  ${}_{\lambda}\hat{R}_{sMPS}$  can be evaluated as

$$ER_s = \lambda \left( 1 - \frac{1}{\beta} \right)^{-s}, \quad \beta > 1, \quad \text{and} \quad E_{\lambda}\hat{R}_{sMPS} = \lambda \left( 1 - \frac{s}{r\beta} \right)^{-r}, \quad \beta > \frac{s}{r}, \quad (7)$$

and are infinite, otherwise. These expected values may be applied to correct the maximum-product-of-spacings predictor. However, a bias correction is not reasonable in general, since finite existence of the expectations depends on the unknown parameter  $\beta$ . It might be an option if there is prior information about  $\beta$  to exceed  $s/r$ .

In Sect. 2.2.2, the maximum-product-of-spacings predictor is applied to construct new prediction intervals for the next future record  $R_{r+1}$  based on previous record values  $R_1, \dots, R_r$ .

## 2.2 Interval prediction of future record values

Based on previous record values, some exact and approximate prediction intervals for future Pareto record values are known from the literature; exact prediction intervals can be deduced from the distributions of the pivot statistics

$$\frac{r-1}{s-r} \frac{\ln R_s - \ln R_r}{\ln R_r - \ln R_1} \sim F(2(s-r), 2(r-1)), \quad (8a)$$

$$\frac{r}{s-r} \frac{\ln R_s - \ln R_r}{\ln R_r - \ln \lambda} \sim F(2(s-r), 2r), \quad r \geq 2, \quad (8b)$$

where  $F(\cdot, \cdot)$  denotes the  $F$  distribution with the respective degrees of freedom [cf. 5, 6, 19, 28].

Due to the relation between  $F$  distributions and  $beta$  distributions, i.e.  $X \sim beta(p, q)$  implies  $\frac{q}{p} \frac{X}{1-X} \sim F(2p, 2q)$ , which in turn implies  $\frac{p}{q} \frac{1-X}{X} \sim F(2q, 2p)$ , the pivot statistics can also be stated as

$$\frac{\ln R_r - \ln R_1}{\ln R_s - \ln R_1} \sim beta(r-1, s-r) \quad \text{and} \quad \frac{\ln R_r - \ln \lambda}{\ln R_s - \ln \lambda} \sim beta(r, s-r), \quad (9)$$

where the density function of  $beta(p, q)$ ,  $p, q > 0$ , is given by  $\frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1}$ ,  $x \in (0, 1)$ , where  $B(\cdot, \cdot)$  denotes the beta function.

Although the results could also be applied for an unknown scale parameter  $\lambda$ , the value of  $\lambda$  is considered known, here, since it represents a lower threshold in the data. Moreover, we examine one-sided prediction intervals for  $R_s, s > r$ , with  $R_r$  as lower bound, too.

In view of the data sets and in contrast to the above references, which consider knowledge of previous record values, only, we plug in estimates of the unknown shape parameter  $\beta$  based on all existing data above threshold  $\lambda$ , whenever necessary.

We supplement known and modified prediction intervals by two constructions based on the maximum-product-of-spacings predictor. In Sect. 3, these prediction intervals for future Pareto record claims are compared in a simulation study by means of coverage probabilities and empirical quantiles of their lengths.

### 2.2.1 Exact prediction intervals

Exact two-sided prediction intervals via pivotals (9) are given by

$$PI_{e1} = \left[ R_1 \exp \left\{ \frac{\ln R_r - \ln R_1}{\beta_{1-\alpha/2}(r-1, s-r)} \right\}, R_1 \exp \left\{ \frac{\ln R_r - \ln R_1}{\beta_{\alpha/2}(r-1, s-r)} \right\} \right] \text{ and} \quad (10)$$

$$PI_{e2} = \left[ \lambda \exp \left\{ \frac{\ln R_r - \ln \lambda}{\beta_{1-\alpha/2}(r, s-r)} \right\}, \lambda \exp \left\{ \frac{\ln R_r - \ln \lambda}{\beta_{\alpha/2}(r, s-r)} \right\} \right], \quad (11)$$

where  $\beta_\alpha(\cdot, \cdot)$  denotes an  $\alpha$ -quantile of the respective *beta* distribution. In the particular case  $s = r + 1$ , we find  $\beta_\alpha(r-1, 1) = \alpha^{1/(r-1)}$ ,  $\alpha \in (0, 1)$ .

According to (8), these prediction intervals can also be stated as

$$PI_{e1} = \left[ R_r \exp \left\{ \frac{s-r}{r-1} (\ln R_r - \ln R_1) F_{\alpha/2}(2(s-r), 2(r-1)) \right\}, \right. \\ \left. R_r \exp \left\{ \frac{s-r}{r-1} (\ln R_r - \ln R_1) F_{1-\alpha/2}(2(s-r), 2(r-1)) \right\} \right] \text{ and} \quad (12)$$

$$PI_{e2} = \left[ R_r \exp \left\{ \frac{s-r}{r} (\ln R_r - \ln \lambda) F_{\alpha/2}(2(s-r), 2r) \right\}, \right. \\ \left. R_r \exp \left\{ \frac{s-r}{r} (\ln R_r - \ln \lambda) F_{1-\alpha/2}(2(s-r), 2r) \right\} \right], \quad (13)$$

where  $F_\alpha(\cdot, \cdot)$  denotes an  $\alpha$ -quantile of the *F* distribution with the respective degrees of freedom; cf. Asgharzadeh et al. [5, for  $s = r + 1$ ] and Empacher et al. [19]. From the two-sided prediction intervals  $PI_{e1}$  and  $PI_{e2}$ , one-sided upper prediction intervals  $PI_{e1}^u$  and  $PI_{e2}^u$  with  $R_r$  as lower bound are obtained by plugging in respective  $\alpha$ - and  $(1 - \alpha)$ -quantiles in the upper bounds of (10), (11) and (12), (13).

## 2.2.2 Approximate prediction intervals

Observing that the statistic

$$2\beta (\ln R_s - \ln R_r) \sim \chi^2 (2(s - r))$$

is a pivot statistic with a chi-squared distribution with  $2(s - r)$  degrees of freedom, the interval

$$\left[ R_r \exp \left\{ \frac{1}{2\beta} \chi_{\alpha/2}^2 (2(s - r)) \right\}, R_r \exp \left\{ \frac{1}{2\beta} \chi_{1-\alpha/2}^2 (2(s - r)) \right\} \right]$$

is an exact prediction interval for  $R_s$ ,  $s > r$ , if  $\beta$  is known, and

$$PI_{a1} = \left[ R_r \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{\alpha/2}^2 (2(s - r)) \right\}, R_r \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{1-\alpha/2}^2 (2(s - r)) \right\} \right] \quad (14)$$

is an approximate prediction interval, if  $\beta$  is replaced by a consistent estimator  $\hat{\beta}$  of  $\beta$ , where  $\chi_{\alpha}^2(p)$  denotes the  $\alpha$ -quantile of a chi-squared distribution with  $p$  degrees of freedom.

**Remark 4** If only record values are available,  $\hat{\beta}$  can be chosen as  $\hat{\beta} = r/(\ln R_r - \ln R_1)$  or  $\hat{\beta} = r/(\ln R_r - \ln \lambda)$ , if  $\lambda$  is known [cf. 19]. However, in view of the real data sets considered in Sect. 4, we plug in maximum likelihood estimates of the unknown shape parameter  $\beta$  based on all claim observations above a given threshold.

In the particular case  $s = r + 1$ , we have  $\chi_{\alpha}^2(2) = -2 \ln(1 - \alpha)$ ,  $\alpha \in (0, 1)$ . For  $s = r + 1$ ,  $PI_{a1}$  simplifies to

$$\left[ R_r (1 - \alpha/2)^{-1/\hat{\beta}}, R_r (\alpha/2)^{-1/\hat{\beta}} \right],$$

which has been noticed in Raqab et al. [28].

The distribution of  $R_s$ ,  $s > r$ , which is the record value to be predicted, can also be a starting point to derive a prediction interval. Since  $2\beta(\ln R_s - \ln \lambda) \sim \chi^2(2s)$ , a consistent estimator  $\hat{\beta}$  of  $\beta$  leads to the approximate prediction interval

$$PI_{a2} = \left[ \lambda \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{\alpha/2}^2 (2s) \right\}, \lambda \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{1-\alpha/2}^2 (2s) \right\} \right],$$

where the quantity  $R_r$  itself does not explicitly appear. However, note that the lower bound can be replaced by

$$\max \left\{ R_r, \lambda \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{\alpha/2}^2 (2s) \right\} \right\}.$$

If the estimator  $\hat{\beta}$  of  $\beta$  is supposed to be based only on record values, then the choice of  $\hat{\beta} = r/(\ln R_r - \ln \lambda)$  leads to a prediction interval shown in Empacher et al. [19].

The following construction of an upper prediction interval is based on the maximum-product-of-spacings predictor  ${}_{\lambda}\hat{R}_{sMPSP}$  in (6) via  $(a \cdot {}_{\lambda}\hat{R}_{sMPSP}, b \cdot {}_{\lambda}\hat{R}_{sMPSP})$  with suitable factors  $a > 0$  and  $b > 0$ . We consider the most important case  $s = r + 1$ , only. For ease of notation, let  ${}_{\lambda}\hat{R}_{(r+1)MPSP} = R^*$ , say.

**Theorem 1** *With the above notations and for a Pareto distribution as in (5), we find:*

$$\begin{aligned} i) \quad a \geq 1 : P(R_{r+1} > aR^*) &= a^{-\beta} \left(1 + \frac{1}{r}\right)^{-r} \\ ii) \quad a < 1 : P(R_{r+1} > aR^*) &= 1 - a^{r\beta} \sum_{j=0}^{r-1} \frac{1}{j!} (-r\beta \log a)^j \\ &\quad + a^{r\beta} \left(1 + \frac{1}{r}\right)^{-r} \sum_{j=0}^{r-1} \frac{1}{j!} (-(r+1)\beta \log a)^j \end{aligned}$$

**Proof** Let  $a \geq 1$ . With the notation  $\tilde{x} = a\lambda(\frac{x}{\lambda})^{\frac{r+1}{r}}$  we obtain by using the Markov property of record values (4):

$$\begin{aligned} P(R_{r+1} > aR^*) &= \int_{\lambda}^{\infty} P(R_{r+1} > \tilde{x} | R_r = x) dP^{R_r}(x) = \int_{\lambda}^{\infty} \frac{1 - F(\tilde{x})}{1 - F(x)} dP^{R_r}(x) \\ &= \frac{\beta^r}{\lambda a^{\beta}(r-1)!} \int_{\lambda}^{\infty} \left(\frac{\lambda}{x}\right)^{\beta/r+\beta+1} \left(\log \frac{x}{\lambda}\right)^{r-1} dx \\ &= \frac{\beta^r}{a^{\beta}(r-1)!} \int_1^{\infty} z^{-(\beta/r+\beta+1)} (\log z)^{r-1} dz \\ &= a^{-\beta} \left(1 + \frac{1}{r}\right)^{-r} \end{aligned}$$

For  $a < 1$ , the integrand  $P(R_{r+1} > \tilde{x} | R_r = x) = 1$ , if  $\tilde{x} < x$ ; hence, in the interval  $(\lambda, \frac{\lambda}{a^r})$ , the integrand equals 1. Thus,

$$\begin{aligned} P(R_{r+1} > aR^*) &= \int_{\lambda}^{\lambda/a^r} dP^{R_r}(x) + \frac{\beta^r}{a^{\beta}(r-1)!} \int_{1/a^r}^{\infty} z^{-(\beta/r+\beta+1)} (\log z)^{r-1} dz \\ &= 1 - a^{r\beta} \sum_{j=0}^{r-1} \frac{1}{j!} (-r\beta \log a)^j \\ &\quad + \frac{1}{a^{\beta}(r-1)!(1 + \frac{1}{r})^r} \Gamma(r, -(r+1)\beta \log a), \end{aligned}$$

which directly leads to the assertion by evaluating the incomplete gamma function.  $\square$

It is worth noticing that the probability  $P(R_{r+1} > aR^*)$  turns out to be independent of  $\lambda$ .

For  $a = 1$ , the probability that  $R_{r+1}$  exceeds  $R^* = {}_{\lambda}\hat{R}_{(r+1)MPS}$  is obtained, namely

$$P(R_{r+1} > R^*) = \left(1 + \frac{1}{r}\right)^{-r} \longrightarrow \frac{1}{e}, \quad r \rightarrow \infty. \quad (15)$$

In terms of expectation, the quantities  $ER^*$  and  $ER_{r+1}$  are ordered, as the following lemma shows.

**Lemma 2** *Let  $\beta > 1 + \frac{1}{r}$ . Then  $ER^* > ER_{r+1}$ .*

**Proof** From (7) we find

$$ER^* \geq ER_{r+1} \iff \left(1 + \frac{1}{r\beta - r - 1}\right)^r \geq \frac{\beta}{\beta - 1},$$

where this inequality is seen to be valid by the Bernoulli-inequality:

$$\left(1 + \frac{1}{r\beta - r - 1}\right)^r \geq 1 + r \frac{1}{r\beta - r - 1} = \frac{\beta - 1/r}{\beta - 1/r - 1} > \frac{\beta}{\beta - 1}.$$

$\square$

The inequality in Lemma 2 is even valid for  $\beta > 1$ , but trivial for  $1 < \beta \leq 1 + 1/r$ , since then  $ER^* = \infty$  and  $ER_{r+1} < \infty$ .

**Remark 5**

- (i) In Theorem 1 i), the probability tends to 0, if  $a$  tends to infinity, and it is bounded from above by  $(1 + \frac{1}{r})^{-r}$ . Hence, for  $\alpha < 1/e \approx 0.37$ , an upper  $(1 - \alpha)$ -prediction interval is obtained as

$$[R_r, bR^*],$$

where  $b = b(\beta) = (\alpha(1 + \frac{1}{r})^r)^{-1/\beta}$ , when  $\beta$  is supposed to be known, and it becomes an approximate  $(1 - \alpha)$ -prediction interval, if  $\beta$  is replaced by a consistent estimator, i.e.

$$PI_{a3}^u = \left[ R_r, \lambda b(\hat{\beta}) \left( \frac{R_r}{\lambda} \right)^{1+1/r} \right].$$

- (ii) If an approximate two-sided equal tail prediction interval of the form  $(aR^*, bR^*)$  is required, then  $a = a(\hat{\beta})$  has to be determined via Theorem 1 ii) by solving

$$a^{r\hat{\beta}} \sum_{j=0}^{r-1} \frac{1}{j!} (-r\hat{\beta} \log a)^j - \frac{a^{r\hat{\beta}}}{(1 + \frac{1}{r})^r} \sum_{j=0}^{r-1} \frac{1}{j!} (-(r+1)\hat{\beta} \log a)^j = \frac{\alpha}{2}$$

with respect to  $a$  and  $b = b(\hat{\beta}) = (\frac{\alpha}{2}(1 + \frac{1}{r})^r)^{-1/\hat{\beta}}$ , where  $\hat{\beta}$  is a consistent estimator of  $\beta$ . It may happen that  $aR^* < R_r$ . Hence, the lower bound of the prediction interval can be chosen as  $\max\{R_r, a(\hat{\beta})R^*\}$ . Accordingly, an approximate  $(1 - \alpha)$ -prediction interval is given by

$$PI_{a3} = \left[ \max \left\{ R_r, \lambda a(\hat{\beta}) \left( \frac{R_r}{\lambda} \right)^{1+1/r} \right\}, \lambda \left( \frac{\alpha}{2} \left( 1 + \frac{1}{r} \right)^r \right)^{-1/\hat{\beta}} \left( \frac{R_r}{\lambda} \right)^{1+1/r} \right].$$

**Remark 6** In the particular situation of an underlying Pareto distribution as in (5) with known threshold parameter  $\lambda > 0$ , we find that  $X/\lambda \sim \text{Par}(1, \beta)$  iff  $X \sim \text{Par}(\lambda, \beta)$ . Hence, by scale transformation of the data,  $\text{Par}(1, \beta)$  could be chosen without loss of generality. Moreover, the transformed record values  $R_i/\lambda = {}_1R_i$ , say, are distributed as record values based on  $\text{Par}(1, \beta)$ . For a  $(1 - \alpha)$ -prediction interval with lower and upper bounds  $l(R_1, \dots, R_r)$  and  $u(R_1, \dots, R_s)$ , respectively, we conclude:

$$\begin{aligned} 1 - \alpha &= P(R_s \in (l(R_1, \dots, R_r), u(R_1, \dots, R_r))) \\ &= P\left(\frac{R_s}{\lambda} \in \left(\frac{1}{\lambda}l(R_1, \dots, R_r), \frac{1}{\lambda}u(R_1, \dots, R_r)\right)\right) \\ &= P\left({}_1R_s \in \left(\frac{1}{\lambda}l(\lambda \cdot {}_1R_1, \dots, \lambda \cdot {}_1R_r), \frac{1}{\lambda}u(\lambda \cdot {}_1R_1, \dots, \lambda \cdot {}_1R_r)\right)\right). \end{aligned}$$

If the latter bounds are independent of the scale parameter  $\lambda$ , then the respective prediction interval has an invariance property against scale normalization. This property is shared by the exact prediction intervals  $PI_{e1}$  and  $PI_{e2}$  as well as by the approximate prediction intervals  $PI_{a1}$ ,  $PI_{a2}$  and  $PI_{a3}$ .

An alternative construction of a prediction interval for the future record value  $R_{r+1}$  based on the maximum-product-of-spacings predictor  $R^*$  is of the form  $[(R^*)^a, (R^*)^b]$  with suitable exponents  $a > 0$  and  $b > 0$ , where for simplicity, we restrict ourselves to  $\lambda \geq 1$ .

**Theorem 3** With the above notations and for a Pareto distribution as in (5), we find for  $\lambda \geq 1$

- i)  $a \geq 1$  :  $P(R_{r+1} > (R^*)^a) = \lambda^{-\beta(a-1)} \left(a \frac{r+1}{r}\right)^{-r}$ ,
- ii)  $\frac{r}{r+1} < a < 1$  :
 
$$P(R_{r+1} > (R^*)^a) = 1 - \tilde{\lambda}^\beta \sum_{j=0}^{r-1} \frac{1}{j!} (-\log \tilde{\lambda}^\beta)^j$$

$$+ \left(a \frac{r+1}{r}\right)^{-r} \tilde{\lambda}^\beta \sum_{j=0}^{r-1} \frac{1}{j!} \left(-a\beta \frac{r+1}{r} \log \tilde{\lambda}\right)^j, \text{ where } \tilde{\lambda} = \lambda^{\frac{r(a-1)}{r(a-1)+a}},$$
- iii)  $a \leq \frac{r}{r+1}$  :  $P(R_{r+1} > (R^*)^a) = 1$ .

**Proof** Let  $a \geq 1$ . With the notation  $\tilde{x} = (\lambda(\frac{x}{\lambda})^{\frac{r+1}{r}})^a$  and by using the Markov property of record values (4), we obtain as in the proof of Theorem 1

$$\begin{aligned} P(R_{r+1} > (R^*)^a) &= \int_{\lambda}^{\infty} P(R_{r+1} > \tilde{x} | R_r = x) dP^{R_r}(x) = \int_{\lambda}^{\infty} \frac{1 - F(\tilde{x})}{1 - F(x)} dP^{R_r}(x) \\ &= \frac{\beta^r}{(r-1)! \lambda^{a\beta - \beta + 1}} \int_{\lambda}^{\infty} \left(\frac{\lambda}{x}\right)^{a\beta \frac{r+1}{r} + 1} \left(\log \frac{x}{\lambda}\right)^{r-1} dx \\ &= \lambda^{-\beta(a-1)} \left(a \frac{r+1}{r}\right)^{-r}. \end{aligned}$$

For  $\frac{r}{r+1} < a < 1$ , we have  $\tilde{x} < x$  in the interval  $(\lambda, \lambda')$  with  $\lambda' = \lambda^{\frac{a}{a(r+1)-r}}$  and thus

$$P(R_{r+1} > (R^*)^a) = \int_{\lambda}^{\lambda'} dP^{R_r}(x) + \int_{\lambda'}^{\infty} \frac{1}{\lambda^{\beta(a-1)}} \left(\frac{x}{\lambda}\right)^{\beta} \left(\frac{\lambda}{x}\right)^{a\beta \frac{r+1}{r}} dP^{R_r}(x),$$

which, by plugging in the density function of  $R_r$  and substituting, yields the assertion.

If  $a \leq \frac{r}{r+1}$ , then  $\tilde{x} < x$  and  $P(R_{r+1} > \tilde{x} | R_r = x) = 1$  on  $(\lambda, \infty)$ .  $\square$

In the case  $a = 1$  in Theorem 3, we reproduce equation (15).

The choice of  $\lambda = 1$ , for which  $\lambda = \lambda'$  in the proof above, yields  $P(R_{r+1} > (R^*)^a) = (a \frac{r+1}{r})^{-r}$  for all  $a \in (\frac{r+1}{r}, \infty)$ , which is independent of  $\beta$ . However, in this case  $\lambda = 1$ , the construction of a prediction interval for  $R_{r+1}$  ends up with the exact interval  $PI_{e2}$ , since  $a = \frac{r}{r+1} (1 - \frac{\alpha}{2})^{-1/r}$  and  $b = \frac{r}{r+1} (\frac{\alpha}{2})^{-1/r}$ .

**Remark 7** (i) For  $a \geq 1$ , the expression in Theorem 3 i) is decreasing in  $a$ , it is bounded from above by  $(\frac{r+1}{r})^{-r}$  and it tends to 0, if  $a \rightarrow \infty$ . Hence, for small values of  $\alpha$ , an upper  $(1 - \alpha)$ -prediction interval is obtained as

$$\left[ R_r, (R^*)^b \right],$$

where  $b = b(\beta)$  is uniquely determined by  $\lambda^{-\beta(b-1)} (b \frac{r+1}{r})^{-r} = \alpha$ . It becomes an approximate  $(1 - \alpha)$ -prediction interval, if  $\beta$  is replaced by a consistent estimator, i.e.

$$PI_{a4}^u = \left[ R_r, \left( \lambda \left( \frac{R_r}{\lambda} \right)^{\frac{r+1}{r}} \right)^{b(\hat{\beta})} \right].$$

In contrast to all previous prediction intervals, the choice  $\lambda = 1$  cannot be made without loss of generality because of the particular derivation of  $PI_{a4}^u$  (see Remark 6).



- (ii) By means of Theorem 3 ii), an approximate two-sided prediction interval of the form  $[(R^*)^a, (R^*)^b]$  can be constructed along the lines of Remark 5(ii).

### 3 Simulation study regarding prediction intervals

In view of applying the prediction intervals to insurance data, we consider upper prediction intervals. We summarize those upper  $(1 - \alpha)$ -prediction intervals as introduced in Sect. 2:

$$\begin{aligned} PI_{e1}^u &= \left[ R_r, R_1 \exp \left\{ \frac{\ln R_r - \ln R_1}{\beta_\alpha(r-1, s-r)} \right\} \right], & PI_{e2}^u &= \left[ R_r, \lambda \exp \left\{ \frac{\ln R_r - \ln \lambda}{\beta_\alpha(r, s-r)} \right\} \right], \\ PI_{a1}^u &= \left[ R_r, R_r \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{1-\alpha}^2(2(s-r)) \right\} \right], & PI_{a2}^u &= \left[ R_r, \lambda \exp \left\{ \frac{1}{2\hat{\beta}} \chi_{1-\alpha}^2(2s) \right\} \right], \\ PI_{a3}^u &= \left[ R_r, \lambda \left( \alpha \left( 1 + \frac{1}{r} \right)^r \right)^{-1/\hat{\beta}} \left( \frac{R_r}{\lambda} \right)^{1+1/r} \right], \\ PI_{a4}^u &= \left[ R_r, \left( \lambda \left( \frac{R_r}{\lambda} \right)^{\frac{r+1}{r}} \right)^{b(\hat{\beta})} \right]. \end{aligned}$$

In  $PI_{a4}^u$ , the exponent  $b(\hat{\beta})$  is obtained by solving  $\lambda^{-\hat{\beta}(b-1)}(b\frac{r+1}{r})^{-r} = \alpha$  in  $b$ . Throughout, the estimator  $\hat{\beta}$  of  $\beta$  is chosen to be based on all data exceeding threshold  $\lambda$  that we observe up to the occurrence of the last record.

Our aims in the following simulation studies are to evaluate the performance of the presented prediction intervals, to compare them regarding coverage and length and, according to those results, to identify preferable prediction intervals, which will be used in Sect. 4.

In the following study, 10,000 sequences of iid Pareto random variables with  $\lambda = 2$  and  $\beta = 1.3$  were simulated by using the statistical software R. In each run, the termination criterion is the observation of the 8-th record. The chosen parameters correspond to parameter values from the global flood losses data set studied in Sect. 4.2. In the Danish fire data set, cf. Section 4.1, the shape parameter is close to 1.3 as well. The scale parameter  $\lambda$  is considered to be known. A change of this scale parameter does not yield different results for the prediction intervals except for  $PI_{a4}^u$  because of the invariance property against scale normalization. Based on the first  $r = 2, \dots, 7$  simulated record values, the next record value ( $s = r + 1$ ) is predicted, successively.

First, the impact of different ways of estimating  $\beta$  is studied. When having knowledge only of the record values, the maximum-likelihood (ML) estimator

$$\hat{\beta} = \frac{r}{\ln(R_r) - \ln(\lambda)}$$

is used. In our simulation setting, we know about the complete data and we can use successive maximum likelihood estimators of these data

$$\hat{\beta}_{L(r)} = \frac{L(r)}{\sum_{i=1}^{L(r)} \ln(X_i) - L(r) \log(\lambda)}, \quad r = 1, \dots, 7,$$

with  $L(r)$  as in (1). Table 2 shows that the usage of the underlying data compared to the estimation based only on record values increases the performances of the approximate prediction intervals considerably in terms of coverage.

In Table 2, it can also be seen that the percentage of coverage of the interval  $PI_{a2}^u$  tends to be too conservative when estimating based on record values. The coverage of  $PI_{a4}^u$  is close to the prediction level if  $r$  is small, when the ML estimate  $\hat{\beta}_{L(r)}$  is applied. It shows the best behavior of the approximate intervals in terms of coverage and is comparable to  $PI_{a1}^u$  and  $PI_{a3}^u$  for  $r \geq 4$ .

In Fig. 2, the lengths of the exact prediction intervals  $PI_{e1}^u$  and  $PI_{e2}^u$  are shown by means of box plots. In this visualization, the upper whiskers indicate largest interval lengths that are smaller than the quartile plus 1.5 times the interquartile range. The case  $r = 2$  with a large box (the upper quartile exceeds 200,000) is omitted to better compare the other boxes. Moreover, a prediction based on  $r = 2$  record observations is not reasonable in practice. It can be seen from Fig. 2, that  $PI_{e1}^u$  is generally larger than  $PI_{e2}^u$ . Among both exact prediction intervals, we recommend the usage of  $PI_{e2}^u$  because of this property. The box plots of both considered exact prediction intervals indicate right skewed empirical distributions with an increasing effect as  $r$  increases ( $r \geq 4$ ), and the interquartile range increases in  $r$ . Note that the simulation study shows quite similar results for repeated simulations (10,000 runs each).

In a next step, the lengths of the approximate prediction intervals are compared in Fig. 3. Again, the simulation study shows quite similar results for repeated simulations. Although the outliers and extremes are different for different repetitions, the resulting box plots differ just marginally. The interquartile ranges of the interval lengths increase as  $r$  increases. The interval  $PI_{a2}^u$  does not show a reasonable behavior as negative lengths are observed. This can happen because the interval does not use the knowledge about the last observed record value. Furthermore, it has the largest median length. Among the remaining approximate intervals  $PI_{a1}^u$  has the smallest lengths and  $PI_{a4}^u$  has the largest lengths. Moreover, the box plots indicate right skewed empirical distribution functions of the lengths of  $PI_{a1}^u$ ,  $PI_{a3}^u$  and  $PI_{a4}^u$  as in the case of the exact prediction intervals.

In Fig. 2 and Fig. 3, the medians of the interval lengths except for  $PI_{a2}^u$  show a usable behavior in the considered cases. Additionally, in Table 3, the median and upper quartile of the lengths of all six prediction intervals are listed. Repeated simulations (with 10,000 sequences, each) yield comparable results.

$PI_{a1}^u$  and  $PI_{a3}^u$  have the smallest medians and upper quartiles.  $PI_{a2}^u$  has the disadvantage of having very large medians and upper quartiles.

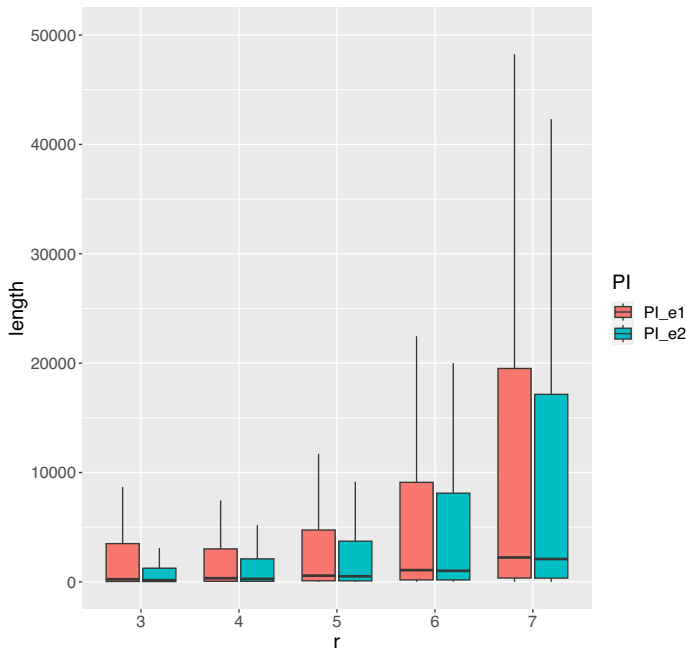
In conclusion, we observe that  $PI_{e2}^u$  outperforms  $PI_{e1}^u$  with respect to interval length.  $PI_{a4}^u$  performs best with respect to coverage along with a reasonable length when the number of observed records is small. For larger numbers of observed records,  $PI_{a1}^u$  and  $PI_{a3}^u$  have smallest lengths followed by  $PI_{a4}^u$ .  $PI_{a2}^u$  does not show a useful

**Table 2** Percentages of coverage of different approximate 90%-prediction intervals for  $s = r + 1$  based on a simulation of 10,000 sequences of Pareto data with (true) parameters  $\lambda = 2$ ,  $\beta = 1.3$

interval based on $r$ records	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
$PI_{a1}^u, \hat{\beta}$ based on records	0.7826	0.8178	0.8427	0.8492	0.8582	0.8642
$PI_{a1}^u, \hat{\beta}_{L(r)}$ based on underlying data	0.8146	0.861	0.8844	0.8888	0.8986	0.8981
$PI_{a2}^u, \hat{\beta}$ based on records	0.8554	0.9122	0.9383	0.9532	0.9657	0.9732
$PI_{a2}^u, \hat{\beta}_{L(r)}$ based on underlying data	0.8645	0.8963	0.9025	0.9024	0.9004	0.8980
$PI_{a3}^u, \hat{\beta}$ based on records	0.7988	0.8314	0.8520	0.8587	0.8642	0.8715
$PI_{a3}^u, \hat{\beta}_{L(r)}$ based on underlying data	0.8242	0.8660	0.8858	0.8901	0.8956	0.8984
$PI_{a4}^u, \hat{\beta}$ based on records	0.7722	0.8404	0.8691	0.8777	0.8842	0.8903
$PI_{a4}^u, \hat{\beta}_{L(r)}$ based on underlying data	0.8581	0.8878	0.8957	0.8980	0.8981	0.9005

**Table 3** Median and upper quartile of the lengths of the different 90%-prediction intervals for  $s = r + 1$  based on 10,000 sequences of Pareto data with (true) parameters  $\lambda = 2$  and  $\beta = 1.3$

interval	median							upper quartile						
	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$		$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	
$PI_{e1}^u$	965	239	333	567	1073	2227		204,282	3498	3015	4741	9102	19,512	
$PI_{e2}^u$	113	150	275	510	1013	2089		1412	1252	2107	3721	8110	17,145	
$PI_{a1}^u$	34	85	190	394	790	1678		154	313	642	1353	3136	7209	
$PI_{a2}^u$	68	227	692	1952	5439	14,672		365	752	1721	3819	8641	20,451	
$PI_{a3}^u$	36	86	187	384	783	1670		199	384	809	1687	4014	9202	
$PI_{a4}^u$	60	118	240	471	953	1996		563	819	1625	3136	7211	15,601	



**Fig. 2** Box plots (without outliers) of lengths of the exact 90%-prediction intervals for  $s = r + 1$  based on a simulation of 10,000 sequences of Pareto data with (true) parameters  $\lambda = 2$  and  $\beta = 1.3$

behavior and is therefore omitted in Sect. 4. Hence, the intervals  $PI_{e2}^u$ ,  $PI_{a1}^u$ ,  $PI_{a3}^u$  and  $PI_{a4}^u$  are used for the following data analyses.

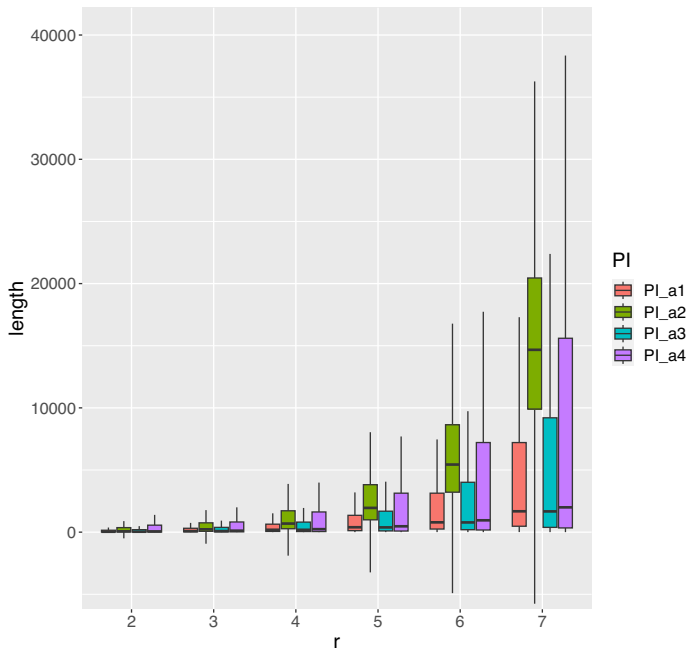
## 4 Real data application

In this section, the developed interval prediction methods are validated and evaluated on real data sets from the insurance industry. Three different data sources are studies for illustrative purposes.

By similarity to the simulation study, we consider 90%-prediction intervals. Further, we estimate the Pareto shape parameter  $\beta$  based on the sequence of all claim resp. loss sizes above the threshold and not by means of the record values, only.

### 4.1 Danish fire losses

First, we consider the well-known Danish insurance data set that has been studied several times in the actuarial literature and that is included in various packages of the software R, see, e.g., the R-cran package ‘evir’ by Pfaff et al. [26]. These data comprise  $n = 2,167$  large fire insurance claims in Denmark over the period 1980 to 1990 collected at Copenhagen Re. The losses are in millions of Danish Krone and have been adjusted to reflect values of the year 1985. All losses are of size one million or

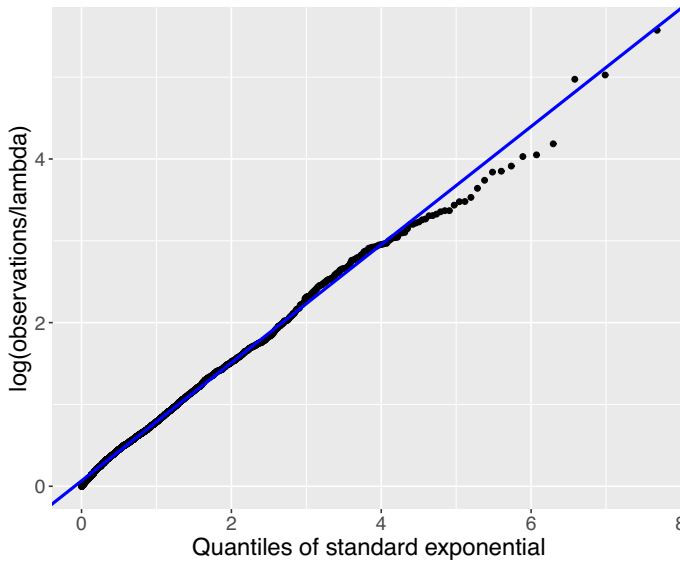


**Fig. 3** Box plots (without outliers) of lengths of the approximate 90%-prediction intervals for  $s = r + 1$  based on a simulation of 10,000 sequences of Pareto data with (true) parameters  $\lambda = 2$  and  $\beta = 1.3$

larger. A diagnostic analysis of the Danish data set was initially performed by McNeil [23] followed by a thorough discussion by Resnick [30]. In Fig. 4, the Pareto quantile-quantile (Q-Q) plot of all the data over one million indicates already a remarkable straight line, as noted also by Resnick [30]. Thus, we use  $\lambda = 1$  as threshold which leads to an ML estimate of  $\hat{\beta}_n = 1.27$  for the Pareto shape parameter.

In Table 4, the sequence of record values  $R_r$ ,  $r = 2, \dots, 7$ , extracted from this data set is shown in the second column. Obviously, there are only a few record values observed, namely seven, and the last record observation is approximately ten times the second last record observation, which may lead to prediction problems [see also 33]. In the third column of Table 4, the corresponding point predictor  ${}_{\lambda}\hat{R}_{rMPS}$  is given, which is indeed not able to reach the magnitude of the last record value. The situation for the upper prediction intervals is different. The fourth column of Table 4 contains the lower bound  $R_{r-1}$  of the respective prediction interval for  $R_r$ . To its right, the upper bounds of the exact interval  $PI_{e2}^u$  as well as of the approximate intervals  $PI_{a1}^u$  and  $PI_{a3}^u$  are given. Note that a column for  $PI_{a4}^u$  can be omitted, since for the particular value  $\lambda = 1$  the interval  $PI_{a4}^u$  coincides with the exact one  $PI_{e2}^u$ .

Since the approximate upper interval bounds of  $PI_{a1}^u$  and  $PI_{a3}^u$  strongly depend on the Pareto shape parameter  $\beta$ , we underline the importance of a proper estimate for  $\beta$ . For illustration, we compute ML estimates based on the sequence of data in two ways. In a retrospective view, we estimate  $\beta$  based on the total of  $n = 2,167$  observation above  $\lambda$  and obtain the estimate  $\hat{\beta}_n = 1.27$ ; the corresponding estimated upper interval



**Fig. 4** Pareto Q-Q plot for  $n = 2,167$  Danish fire losses above the threshold  $\lambda = 1$  and regression line

**Table 4** Prediction of the successive next record Danish fire loss. The ML estimate  $\hat{\beta}_n = 1.27$  is computed based  $n = 2,167$  loss sizes above the threshold  $\lambda = 1$

$r$	$R_r$	$\lambda \hat{R}_{rMPS}$	$R_{r-1}$	$PI_{e2}^u$	$\hat{\beta}_n$		$\hat{\beta}_{L(r-1)}$		$\hat{\beta}_{L(r-1)}$
					$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a1}^u$	$PI_{a3}^u$	
2	2.09	2.84	1.68	183.14	10.31	10.06	5.59	6.56	1.92
3	4.61	3.03	2.09	10.35	12.82	9.80	8.93	7.75	1.59
4	8.73	7.68	4.61	26.93	28.24	23.83	27.98	23.70	1.28
5	11.37	15.00	8.73	47.09	53.42	45.49	90.01	62.60	0.99
6	26.21	18.50	11.37	47.16	69.65	55.27	152.86	88.87	0.89
7	263.25	45.18	26.21	120.82	160.51	133.60	456.86	249.82	0.81

bounds of  $PI_{a1}^u$  and  $PI_{a3}^u$  are given in the sixth and seventh column of Table 4. In a practical ongoing view, we estimate  $\beta$  based on the successively available data until the index  $L(r-1)$ , which is the number of observations until the previous record value  $R_{r-1}$  has been observed; the respective ML estimates  $\hat{\beta}_{L(r-1)}$  are shown in the last column of Table 4 and the corresponding estimated upper interval bounds of  $PI_{a1}^u$  and  $PI_{a3}^u$  to its left. Both views reflect different levels of information and may have their justifications in practice. Here, the information level up to the previous record is the view that manages to capture the magnitude of all records, including the last one, via interval  $PI_{a1}^u$ , which is based on  $\hat{\beta}_{L(r-1)}$ .

Regarding the Danish fire losses, we are faced with the challenge that an extraordinary high record value that is early observed in a sequence of data leads to only a few record observations in total. We therefore consider the sequences of second

**Table 5** Prediction of the successive next 2nd record Danish fire loss. The ML estimate  $\hat{\beta}_n = 1.27$  is computed based on  $n = 2,167$  loss sizes above the threshold  $\lambda = 1$

$r$	$R_r^{(2)}$	$\lambda \hat{R}_{rMPS}^{(2)}$	$R_{r-1}^{(2)}$	$PI_{e2}^u$	$\hat{\beta}_n$		$\hat{\beta}_{L^{(2)}(r-1)}$		$\hat{\beta}_{L^{(2)}(r-1)}$
					$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a1}^u$	$PI_{a3}^u$	
2	1.73	2.84	1.68	183.14	4.17	5.34	3.48	4.71	1.59
3	1.78	2.28	1.73	5.69	4.29	4.10	3.47	3.58	1.66
4	2.09	2.16	1.78	3.46	4.40	3.80	3.54	3.31	1.68
5	4.61	2.52	2.09	3.72	5.18	4.39	5.16	4.37	1.28
6	7.90	6.26	4.61	11.28	11.41	10.82	14.81	12.67	0.99
7	8.73	11.15	7.90	20.77	19.55	19.17	30.17	24.85	0.86
8	11.37	11.89	8.73	20.29	21.59	20.37	31.99	25.72	0.89
9	14.12	15.41	11.37	25.59	28.15	26.33	47.49	35.86	0.81
10	17.57	18.95	14.12	30.56	34.94	32.29	60.82	44.73	0.79
11	21.96	23.40	17.57	36.90	43.47	39.80	67.49	51.50	0.86
12	26.21	29.08	21.96	45.09	54.34	49.38	80.80	62.26	0.88
13	34.14	34.42	26.21	52.31	64.87	58.36	97.85	74.16	0.87
14	56.23	44.79	34.14	67.67	84.48	75.87	115.02	90.79	0.95
15	65.71	74.98	56.23	115.55	139.13	126.87	186.54	150.41	0.96
16	152.41	86.85	65.71	131.58	162.59	146.84	191.55	161.47	1.08

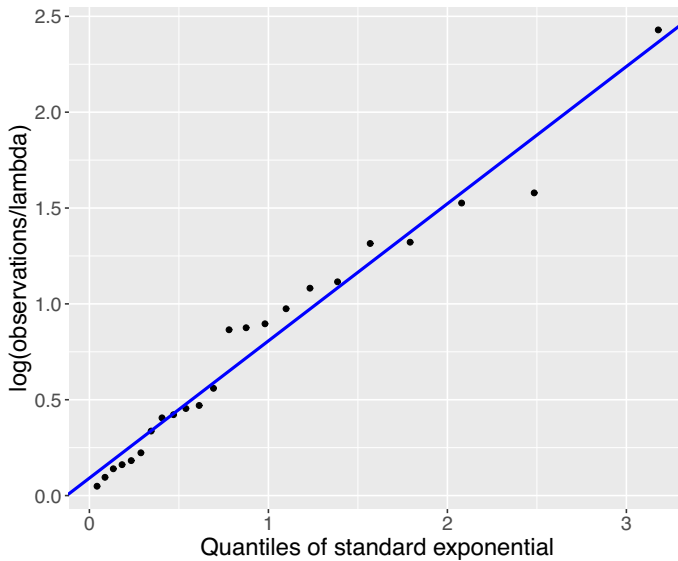
record values, that is  $k = 2$  in (2), which can be expected to show a smoother record development in time. In Table 5, the sequence  $R_r^{(2)}$  of second record values extracted from the Danish data set is shown for  $r = 2, \dots, 16$ . Obviously, the record sample size increases to sixteen observed second record claim sizes. By definition, the first record values, with the exception of the last one, are included in the sequence of second ones. The prediction results in Table 5 show that both the differences between the prediction intervals as well as the interval lengths become smaller compared to first record values in Table 4. Moreover, as can be seen in Table 5, the prediction intervals capture the magnitude of the sequence of second record observations; especially the approximate intervals  $PI_{a1}^u$  and  $PI_{a3}^u$  achieve again to capture even the last second record observation in this sequence when using the ML estimate  $\hat{\beta}_{L(r-1)}$ , as also  $PI_{a1}^u$  does when using the ML estimate  $\hat{\beta}_n$  instead.

To conclude on the Danish fire insurance data set, we observe that in the case of only a few observed record values with an extraordinary last record, the study of successive second or third largest record values provides an option for statistical inference and prediction.

## 4.2 Global flood losses

Next, we consider a data set from the research study 'sigma-1/2021' by the Swiss Re Institute [see 11]. The study shows global annual insured losses for the period 1970 to 2020 resulting from specific natural catastrophe perils. The losses are reported in USD





**Fig. 5** Pareto Q-Q plot for  $n = 23$  global flood losses above the threshold  $\lambda = 2$  and regression line

**Table 6** Original and modified data sequence of  $n = 23$  global flood losses above the threshold  $\lambda = 2$

original data	2.3	4.8	3.0	2.1	2.4	4.8	3.2	7.5
	2.3	4.9	7.5	2.5	2.8	5.3	22.7	3.5
	9.2	3.0	5.9	9.7	2.2	3.2	6.1	
modified data	2.30	4.80	3.00	2.10	2.40	4.75	3.20	7.50
	2.35	4.90	7.45	2.50	2.80	5.30	22.70	3.50
	9.20	3.05	5.90	9.70	2.20	3.15	6.10	

billions and are adjusted to reflect prices of the year 2020. We consider catastrophic annual flood losses, in the following referred to as global flood losses, to fit a Pareto distribution with threshold  $\lambda = 2$  on  $n = 23$  loss sizes. The corresponding Pareto Q-Q plot is given in Fig. 5, indicating a good fit to a straight line, again. Since the original data are reported with only one decimal place, ties have been resolved by alternately subtracting or adding the value 0.05 to identical loss sizes. The original and modified sequences of losses can be seen in Table 6.

The successive records, which have been extracted from the modified data, are listed in Table 7 along with the prediction results. Due to the threshold  $\lambda = 2$ , the upper bounds of the approximate prediction interval  $PI_{a4}^u$  differ from the upper bounds of the exact prediction interval  $PI_{e2}^u$  and are thus reported as well. Although only four successive next record values were observed, the prediction intervals provide reasonable results and are even able to capture the largest successive record that is around three times in magnitude than the previous one. Thereby, the approximate prediction intervals that are calculated by using the successively obtained ML estimates

**Table 7** Prediction of the successive next record global flood loss. The ML estimate  $\hat{\beta}_n = 1.32$  is computed based  $n = 23$  loss sizes above the threshold  $\lambda = 2$ 

$r$	$R_r$	$\lambda \hat{R}_{rMPSP}$	$R_{r-1}$	$\hat{\beta}_n$				$\hat{\beta}_{L(r-1)}$			
				$PI_{e2}^u$	$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	$\hat{\beta}_{L(r-1)}$
2	4.8	2.64	2.3	8.09	13.23	8.99	7.02	3.17	3.31	3.46	7.16
3	7.5	7.44	4.8	31.87	27.61	23.10	24.88	15.45	15.86	20.25	1.97
4	22.7	11.65	7.5	34.49	43.15	34.79	32.80	25.92	25.30	29.29	1.86

**Table 8** Prediction of the successive next 2nd record global flood loss. The ML estimate  $\hat{\beta}_n = 1.32$  is computed based  $n = 23$  loss sizes above the threshold  $\lambda = 2$ 

$r$	$R_r^{(2)}$	$\lambda \hat{R}_{rMPSP}^{(2)}$	$R_{r-1}^{(2)}$	$\hat{\beta}_n$				$\hat{\beta}_{L^{(2)}(r-1)}$			
				$PI_{e2}^u$	$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	$\hat{\beta}_{L^{(2)}(r-1)}$
2	3.00	2.64	2.30	8.09	5.52	4.88	4.83	4.13	3.98	4.11	1.97
3	4.75	3.67	3.00	7.21	7.20	6.48	6.41	5.17	5.23	5.57	2.11
4	4.80	6.34	4.75	12.89	11.39	10.95	11.56	7.70	8.57	9.87	2.38
5	4.90	5.97	4.80	9.49	11.51	10.21	9.55	8.92	8.73	8.98	1.86
6	7.45	5.86	4.90	8.28	11.75	9.94	8.64	9.09	8.51	8.27	1.86
7	7.50	9.28	7.45	13.78	17.87	15.66	14.10	14.99	14.09	13.74	1.65
8	9.20	9.06	7.50	12.55	17.99	15.23	13.02	16.98	14.72	12.95	1.41
9	9.70	11.13	9.20	15.30	22.07	18.67	15.85	21.79	18.53	15.83	1.34

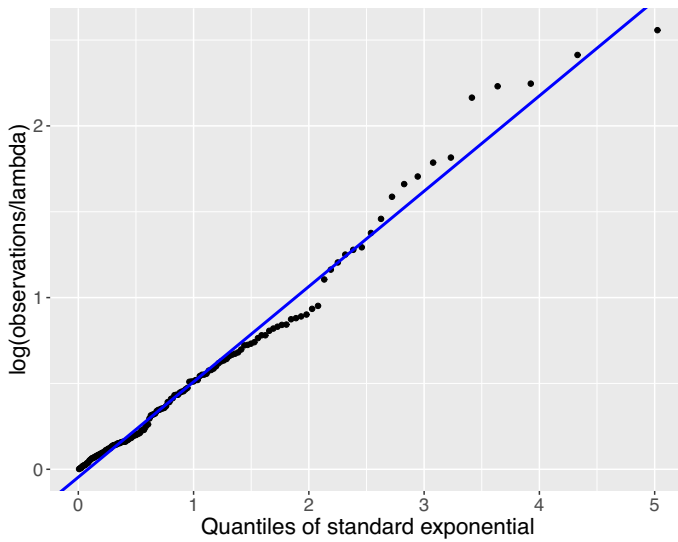
$\hat{\beta}_{L(r-1)}$  give the most meaningful results (for  $r = 3, 4$ ) in terms of smallest interval lengths.

Due to the small sample size of three successively observed record values, we consider additionally the sequence of successive next second record values, even if the prediction results are already very satisfactory. As can be seen in Table 8, the sample size increases to  $r = 9$ . Regarding the prediction results in Table 8, as with the Danish fire losses, we observe with the global flood losses that a larger record sample size increases prediction accuracy. That is, the differences between the prediction intervals as well as the interval lengths become smaller when considering second record values compared to original record values only.

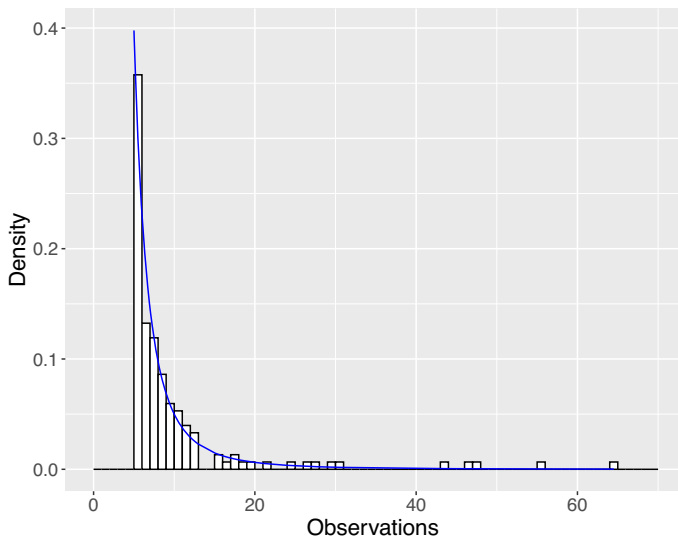
### 4.3 Reinsured catastrophe losses

As a third data source, we were provided with data from an international non-life reinsurer on its worldwide reinsured catastrophe losses per event. We were able to evaluate losses above EUR 5 million and to distinguish between man-made and natural disasters. The losses are reported in million EUR and are both inflation and portfolio adjusted as of the year 2022.

Considering only man-made disaster, the available loss data includes  $n = 151$  observations. In Fig. 6, the corresponding Pareto Q-Q plot shows a reasonable fit to



**Fig. 6** Pareto Q-Q plot for  $n = 151$  reinsured man-made catastrophe losses above the threshold  $\lambda = 5$  and regression line



**Fig. 7** Histogram of  $n = 151$  reinsured man-made catastrophe losses above the threshold  $\lambda = 5$  and pdf of Pareto distribution with ML estimate  $\hat{\beta}_n = 2.00$

a straight line. Thus, we fit a Pareto distribution with threshold  $\lambda = 5$  and obtain an ML estimate  $\hat{\beta}_n = 2.00$  for the Pareto shape parameter. In Fig. 7, the fit is evaluated by means of a histogram compared to the theoretical density function of the Pareto distribution with the aforementioned parameters.

**Table 9** Prediction of the successive next record reinsured man-made catastrophe loss. The ML estimate  $\hat{\beta}_n = 2.00$  is computed based on  $n = 151$  loss sizes above the threshold  $\lambda = 5$

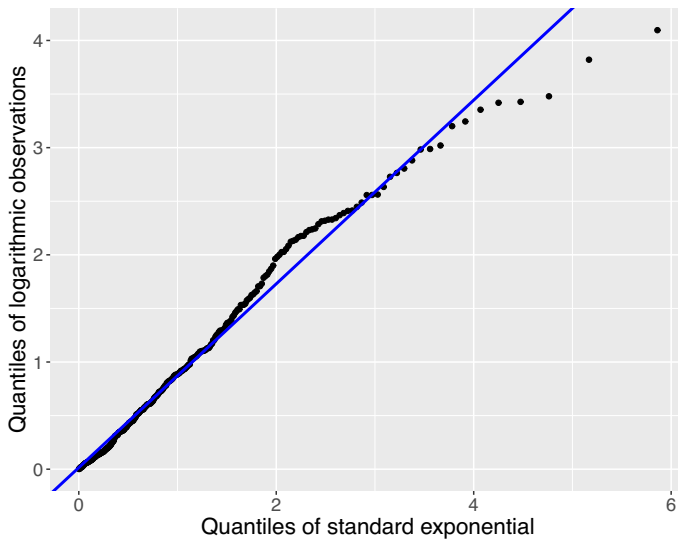
$r$	$R_r$	$\lambda \hat{R}_{rMPSP}$	$R_{r-1}$	$PI_{e2}^u$	$\hat{\beta}_n$			$\hat{\beta}_{L(r-1)}$			$\hat{\beta}_{L(r-1)}$
					$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	
2	7.39	6.63	5.76	20.50	18.23	14.84	14.04	7.97	8.32	8.48	7.09
3	7.72	8.99	7.39	17.22	23.41	18.97	17.39	13.64	13.37	13.63	3.76
4	8.67	8.92	7.72	12.73	24.43	18.33	15.21	14.76	13.38	12.93	3.55
5	10.73	9.95	8.67	13.31	27.45	20.15	15.96	17.01	15.03	14.14	3.42
6	12.73	12.51	10.73	16.78	33.98	25.09	19.60	21.54	19.05	17.77	3.31
7	17.95	14.87	12.73	19.71	40.30	29.64	22.71	26.56	23.10	21.07	3.13
8	46.53	21.55	17.95	29.53	56.82	42.72	33.12	40.48	34.93	31.39	2.83
9	47.26	61.50	46.53	97.92	147.31	121.48	103.58	118.37	106.76	99.98	2.47
10	55.85	60.65	47.26	90.96	149.60	119.46	97.42	124.80	107.38	95.12	2.37
11	64.51	71.09	55.85	104.32	176.79	139.67	111.75	171.99	137.43	111.43	2.05

In Table 9, the successive records  $R_r$ ,  $r = 1, \dots, 11$ , extracted from the data set along with the prediction results can be seen. For the early observed record values, that is  $r = 2, 3$ , the approximate intervals  $PI_{a1}^u$ ,  $PI_{a3}^u$  and  $PI_{a4}^u$  that are calculated by using the successively obtained ML estimates  $\hat{\beta}_{L(r-1)}$  provide the smallest interval length. After that, from  $r = 4$ , the exact interval  $PI_{e2}^u$  is the one with the smallest interval length followed by the approximate intervals  $PI_{a1}^u$ ,  $PI_{a3}^u$  and  $PI_{a4}^u$  that are again calculated with  $\hat{\beta}_{L(r-1)}$ . However, as the number of records increases, the estimation results for the upper interval bounds become closer, especially between the approximate intervals when they are calculated retrospectively using  $\hat{\beta}_n$  and successively using  $\hat{\beta}_{L(r-1)}$ , respectively. Thus, as with the Danish fire and global flood data sets, we observe how a larger record sample size increases prediction accuracy.

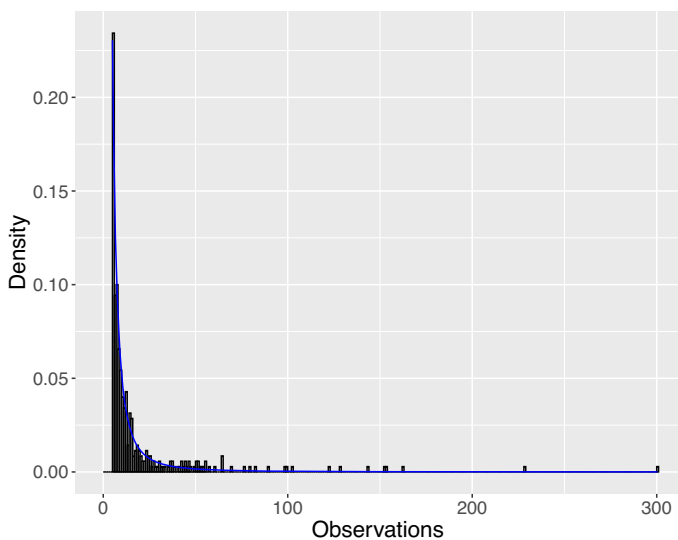
To conclude on the reinsured man-made catastrophe data set, the prediction intervals are able to capture the magnitude of the sequence of record losses in general. There is only one exception with regard to the eighth successive record observation  $R_8$ , which is more than two and a half times in size than the previous record observation, which certainly poses challenges for prediction.

Considering both man-made and natural disasters, the available sequence of loss data comprises  $n = 350$  observations. The Pareto Q-Q plot for these data, referred to as reinsured catastrophe losses in the following, is shown in Fig. 8. The figure indicates that the underlying distribution has a heavier tail compared to the data set that comprises man-made disaster losses only. Nevertheless, fitting a Pareto distribution still seems reasonable. For the threshold  $\lambda = 5$ , we get a ML estimate for the Pareto shape parameter that is  $\hat{\beta}_n = 1.16$ . In Fig. 9, a histogram compared to the theoretical density function of the Pareto distribution with the aforementioned parameters is shown. In particular, the heavier tail of the catastrophe loss data compared to the catastrophe man-made loss data (see Fig. 7) is obvious.

As the extracted records from the sequence of reinsured catastrophe losses comprises a record sample size of only five records, we confine ourselves to showing the



**Fig. 8** Pareto Q-Q plot for  $n = 350$  reinsured catastrophe losses above the threshold  $\lambda = 5$  and regression line



**Fig. 9** Histogram of  $n = 350$  reinsured catastrophe losses above  $\lambda = 5$  and pdf of Pareto distribution with ML estimate  $\hat{\beta}_n = 1.16$

table for the second largest record observations. That is, in Table 10 the sequence  $R_r^{(2)}$  of second largest record reinsured catastrophe losses is shown for  $r = 2, \dots, 11$ . Obviously, the sample size increases to eleven observed second largest record losses. Regarding the early observation  $R_2^{(2)}$ , only the approximate prediction intervals  $PI_{a1}^u$ ,  $PI_{a3}^u$  and  $PI_{a4}^u$  that are calculated with  $\hat{\beta}_n = 1.16$  achieve to capture the magnitude of

**Table 10** Prediction of the successive next 2nd record reinsured catastrophe loss. The ML estimate  $\hat{\beta}_n = 1.16$  is computed based on  $n = 350$  loss sizes above the threshold  $\lambda = 5$

$r$	$R_r^{(2)}$	$\lambda \hat{R}_{rMPSP}^{(2)}$	$R_{r-1}^{(2)}$	$PI_{e2}^u$	$\hat{\beta}_n$			$\hat{\beta}_{L^{(2)}(r-1)}$			$\hat{\beta}_{L^{(2)}(r-1)}$
					$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	$PI_{a1}^u$	$PI_{a3}^u$	$PI_{a4}^u$	
2	8.56	5.23	5.11	6.23	13.84	10.48	9.34	7.06	6.55	6.47	3.57
3	9.86	11.21	8.56	27.42	23.19	21.37	21.61	19.20	18.91	19.71	1.43
4	23.42	12.37	9.86	21.61	26.71	23.06	21.73	21.98	20.42	20.23	1.44
5	27.57	34.45	23.42	77.89	63.40	63.40	67.74	50.52	55.17	61.89	1.50
6	64.84	38.79	27.57	74.84	74.64	70.80	71.38	64.26	64.67	68.16	1.36
7	79.31	99.39	64.84	215.01	175.55	180.36	196.22	151.31	165.01	187.36	1.36
8	128.21	117.70	79.31	232.72	214.71	212.69	221.82	178.13	190.35	210.90	1.42
9	153.95	192.33	128.21	378.32	347.12	346.41	362.81	316.54	328.04	355.12	1.27
10	162.13	225.30	153.95	418.15	416.80	404.74	409.95	413.32	402.75	409.30	1.17
11	228.11	229.59	162.13	399.10	438.95	411.59	399.52	413.46	397.40	395.50	1.23

the record observation. Afterwards, from  $r = 3$  on, the approximate intervals  $PI_{a1}^u$ ,  $PI_{a3}^u$  and  $PI_{a4}^u$  that are calculated with  $\hat{\beta}_{L^{(2)}(r-1)}$  show better results in most cases by means of interval lengths than the retrospectively calculated ones or the exact prediction interval  $PI_{e2}^u$ . However, as the number of second records increases, the estimation results for the upper bounds of the various prediction intervals become close anyway.

Note that we do not cover the record analysis for the data set consisting only of reinsured natural catastrophe losses in this manuscript, since a Pareto fit is not appropriate here. Prediction intervals for ( $k$ -th) records along with recommendations based on other distributions than Pareto, including Weibull and generalized Pareto, are subject to future research. Initial theoretical results in case of a GPD are shown in the next section.

## 5 Modeling with a GPD distribution

Aiming at establishing explicit prediction methods, we were concerned with common Pareto distributions so far. Generalized Pareto distributions may lead to a better fit to data above some threshold; this class of distributions is extensively examined in extreme value theory. GPDs appear as limit distributions of scaled quantities exceeding high thresholds (cf. Beirlant et al. [9, Chapter 5], Embrechts et al. [18, Section 3.4]) and are given by distribution functions of the form

$$F(x) = 1 - \left(1 + \frac{x - \lambda}{\sigma}\right)^{-\beta'} \quad (16)$$

with a positive scale parameter  $\sigma$ . Here we consider the case of a positive shape parameter  $\beta'$  and infinite support  $(\lambda, \infty)$  for some threshold  $\lambda$ , which is non-negative and known in our applications. In the particular case  $\lambda = \sigma$ , we arrive at Pareto

distributions as in (5). In Embrechts et al. [18, Section 6.5], fitting of GPDs via the peaks-over-threshold method is described, and for maximum likelihood estimation of  $\beta'$  and  $\sigma$  we also refer to the above monographs (cf. Beirlant et al. [9, Section 5.3], Embrechts et al. [18, Section 6.5]). Let  $\hat{\beta}'$  and  $\hat{\sigma}$  denote the maximum likelihood estimators of  $\beta'$  and  $\sigma$ , respectively, which are not explicitly available.

For comparison with the Pareto case, we fit GPDs to real data sets considered in Sect. 4. In the situation of an underlying GPD, we introduce a point predictor of the future record value  $R_{r+1}$  based on  $R_1, \dots, R_r$  as well as a prediction interval, which serve as a supplement to the studies in Sect. 4 based on Pareto distributions as well as tools in situations, where a GPD should be used instead of a Pareto distribution.

With the aim to predict  $R_{r+1}$  based on  $R_1, \dots, R_r$  via one of the aforementioned methods, some calculation show that the MLP becomes trivial, i.e.,

$$\hat{R}_{r+1,MLP} = R_r,$$

which is also the case for the MOLP, i.e.,

$$\hat{R}_{r+1,MOLP} = R_r$$

[see 31].

Moreover, the MPSP is not explicitly available [see 32]. Hence, in terms of these point prediction principles and in case of a GPD with unknown scale parameter  $\sigma$ , there is no useful predictor of  $R_{r+1}$  based on the previous record values  $R_1, \dots, R_r$ . However, if  $\sigma$  is assumed to be known, then the MPSP is given by

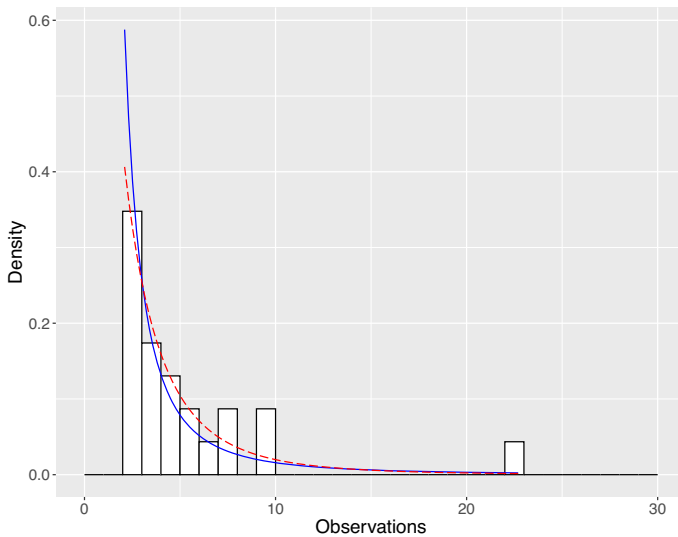
$$\hat{R}_{r+1,MPSP}(\sigma) = \sigma \left( \left( 1 + \frac{R_r - \lambda}{\sigma} \right)^{\frac{r+1}{r}} - 1 \right) + \lambda,$$

where the shape parameter  $\beta'$  drops [see 32].

Thus, replacing  $\sigma$  by a consistent estimator such as  $\hat{\sigma}$ , we obtain the plug-in point predictor  $\hat{R}_{r+1,MPSP}(\hat{\sigma})$  of  $R_{r+1}$ , whose values in real data examples are shown in Tables 11, 12 and 13.

By analogy with the derivation of the exact prediction interval  $PI_{e2}$  in case of an underlying Pareto distribution (see (11)) and noticing that  $R_j$  is distributed as  $F^{-1} \left( 1 - \exp \left( -R_j^E \right) \right)$ ,  $j \in \mathbb{N}$ , where  $R_j^E$  denotes the  $j$ -th record value based on a standard exponential distribution, we find

$$P \left( \frac{\ln \left( \frac{R_r - \lambda}{\sigma} + 1 \right)}{\ln \left( \frac{R_{r+1} - \lambda}{\sigma} + 1 \right)} \leq u \right) = u^r, \quad u \in (0, 1),$$



**Fig. 10** Histogram of  $n = 23$  global flood losses above the threshold  $\lambda = 2$  and pdf of Pareto distribution with ML estimate  $\hat{\beta}_n = 1.32$  (solid blue line) and GPD distribution with ML estimates  $\hat{\beta}'_n = 3.41$ ,  $\hat{\sigma}_n = 7.94$  (dashed red line)

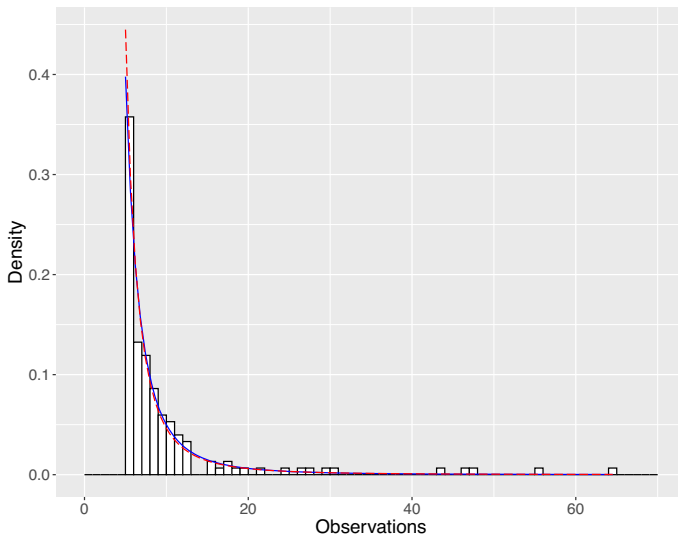
**Table 11** Prediction of the successive next record global flood loss. The ML estimate  $\hat{\sigma}_n = 7.94$  is computed based  $n = 23$  loss sizes above the threshold  $\lambda = 2$

$r$	$R_r$	$\lambda \hat{R}_{rMPSP}$ (Pareto)	$\lambda \hat{R}_{rMPSP}$ (GPD)	$R_{r-1}$	$PI_{e2}^u$ (Pareto)	$PI_{e2}^u$ (GPD)
2	4.8	2.64	2.61	2.3	8.09	5.56
3	7.5	7.44	6.55	4.8	31.87	14.70
4	22.7	11.65	10.08	7.5	34.49	18.73

**Table 12** Prediction of the successive next 2nd record global flood loss. The ML estimate  $\hat{\sigma}_n = 7.94$  is computed based  $n = 23$  loss sizes above the threshold  $\lambda = 2$

$r$	$R_r^{(2)}$	$\lambda \hat{R}_{rMPSP}^{(2)}$ (Pareto)	$\lambda \hat{R}_{rMPSP}^{(2)}$ (GPD)	$R_{r-1}$	$PI_{e2}^u$ (Pareto)	$PI_{e2}^u$ (GPD)
2	3.00	2.64	2.61	2.30	8.09	5.56
3	4.75	3.67	3.55	3.00	7.21	5.61
4	4.80	6.34	5.86	4.75	12.89	9.13
5	4.90	5.97	5.64	4.80	9.49	7.65
6	7.45	5.86	5.60	4.90	8.28	7.06
7	7.50	9.28	8.67	7.45	13.78	11.16
8	9.20	9.06	8.55	7.50	12.55	10.56
9	9.70	11.13	10.47	9.20	15.30	12.84





**Fig. 11** Histogram of  $n = 151$  reinsured man-made catastrophe losses above the threshold  $\lambda = 5$  and pdf of Pareto distribution with ML estimate  $\hat{\beta}_n = 2.00$  (solid blue line) and GPD distribution with ML estimates  $\hat{\beta}'_n = 1.64$ ,  $\hat{\sigma}_n = 3.66$  (dashed red line)

where  $\sigma$  is supposed to be known. Let  $u_\alpha$  denote the respective  $\alpha$ -quantile. Then,

$$PI_{e2}^u(GPD) = \left[ R_r, \sigma \left( \exp \left( \frac{\ln \left( \frac{R_r - \lambda}{\sigma} + 1 \right)}{u_\alpha} \right) - 1 \right) + \lambda \right]$$

is an exact upper  $(1 - \alpha)$ -prediction interval for  $R_{r+1}$ . By plugging in a consistent estimator  $\hat{\sigma}$  of  $\sigma$  again, we obtain an approximate upper  $(1 - \alpha)$ -prediction interval for  $R_{r+1}$ .

In practice, we expect that the class of GPDs with an additional parameter has a better fit to real data than the subclass of Pareto distributions. The histogram in Fig. 11 compares the fit of the Pareto and GPD to the data of global flood losses from Sect. 4. It turns out that the fit of the GPD is only slightly better than the fit of the Pareto distribution. Additionally, the prediction results for record flood losses in Table 11 and the prediction of the 2nd records in Table 12 show no superiority of the GPD prediction compared to the Pareto methods. In fact, the fitted GPD has a lighter tail than the fitted Pareto distribution. The GPD prediction procedure tends to yield smaller upper bounds than the Pareto procedure and is not able to cover all observed records in this example.

The data set of reinsured man-made catastrophe losses with 151 observations is much larger than the flood losses data. The histogram in Fig. 11 shows that the fit of GPD as in (16) and Pareto distribution as in (5) is similar. Therefore, we can expect the prediction results to be comparable as well, which can be observed in Table 13. In this

**Table 13** Prediction of the successive next record reinsured man-made catastrophe loss. The ML estimate  $\hat{\sigma}_n = 3.66$  is computed based  $n = 151$  loss sizes above the threshold  $\lambda = 5$ 

$r$	$R_r$	$\lambda \hat{R}_{rMPSP}$ (Pareto)	$\lambda \hat{R}_{rMPSP}$ (GPD)	$R_{r-1}$	$PI_{e2}^u$ (Pareto)	$PI_{e2}^u$ (GPD)
2	7.39	6.63	6.67	5.76	20.50	25.35
3	7.72	8.99	9.12	7.39	17.22	19.30
4	8.67	8.92	9.01	7.72	12.73	13.44
5	10.73	9.95	10.06	8.67	13.31	13.92
6	12.73	12.51	12.68	10.73	16.78	17.64
7	17.95	14.87	15.10	12.73	19.71	20.71
8	46.53	21.55	21.96	17.95	29.53	31.27
9	47.26	61.50	63.21	46.53	97.92	105.82
10	55.85	60.65	62.15	47.26	90.96	97.31
11	64.51	71.09	72.74	55.85	104.32	111.01

setting, the GPD methods tend to give larger predictors. Nevertheless, both prediction intervals are not able to cover the large jump between the seventh and eighth record.

Other approximate prediction intervals can be derived similar to  $PI_{a1}^u$  and  $PI_{a2}^u$  in Sect. 2.2.2, where parameter estimates of  $\beta'$  and  $\sigma$  are needed. This will be examined in a subsequent manuscript along with comparisons via simulation and recommendations for their use. Prediction intervals such as  $PI_{a3}^u$  and  $PI_{a4}^u$  by making use of the approximate MPSP  $\hat{R}_{r+1,MPSP}$  do not seem to be tractable. The approach to obtain an approximate MPSP and prediction intervals via pivotal statistics may also be applied to other underlying distributions such as Burr XII or Fréchet. However, the derivation of more elaborate prediction intervals relies on the particular form of the distribution function  $F$ .

For general classes of distributions, the above methods for prediction intervals for record values based on quantile functions are not usable, in general. In heavy-tail setups, distributions with regularly varying upper tails are of particular interest. In this case the distribution function  $F$  is given by  $F(x) = 1 - x^{-\beta} L(x)$  with shape parameter  $\beta > 0$ , where  $L$  is a regularly or slowly varying function at infinity. Feuerverger and Hall [20] consider estimation of the shape parameter and Berred [10] studies prediction of future record values in terms of approximate tolerance regions based on an estimated shape parameter, where approximation is with regard to an increasing number of record values. In our situation and commonly, the number of observed record values is fairly small. Nevertheless, the latter approach enables to come up with approximate prediction intervals for general classes of distributions.

## 6 Conclusion

The model of upper record values with an underlying Pareto distribution is shown to be appropriate to predict future record claims or losses based on previously observed record values in a sequence of claims over time. A simulation study focusing on

coverage frequencies and quantiles of lengths leads to selected prediction intervals that can be recommended for practical use, if the assumption of a Pareto distribution appears to be reasonable. These prediction intervals are able to capture the magnitude of future record claims well even if they are applied in the presence of only a small number of observed records. In case of just a few record values being available, the prediction of future second or third largest claims via the model of  $k$ -th record values provides an option for statistical analysis. Statistical prediction of future record values strongly depends on an accurate fit of the underlying distribution. Prediction intervals based on other distributions than Pareto are in progress; an initial result for a generalized Pareto distribution is shown.

**Acknowledgements** The authors are grateful to the associate editor and the reviewers for their careful reading and their constructive criticism, which led to an improved manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Availability of data and materials** The data set considered in Sect. 4.1 is included in various packages of the statistical software R, see, e.g., the R-cran package ‘evir’ by Pfaff et al. [26, <https://cran.r-project.org/web/packages/evir/evir.pdf>]. The data set analyzed in Sect. 4.2 is from the research study ‘sigma-1/2021’ by the Swiss Re Institute [see 11, <https://www.swissre.com/institute/research/sigma-research/sigma-2021-01.html>]. Due to a non-disclosure agreement, the data sets of Sect. 4.3 are not publicly available.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Ahmadi J, Doostparast M (2006) Bayesian estimation and prediction for some life distributions based on record values. *Stat Pap* 47:373–392
2. Ahmadi J, Jafari Jozani M, Marchand E et al (2009) Prediction of  $k$ -records from a general class of distributions under balanced type loss functions. *Metrika* 70(1):19–33
3. Albrecher H, Araujo-Acuna JC, Beirlant J (2021) Tempered Pareto-type modelling using Weibull distributions. *ASTIN Bull* 51(2):509–538
4. Arnold BC, Balakrishnan N, Nagaraja HN (1998) *Records*. Wiley, New York
5. Asgharzadeh A, Abdi M, Kus C (2011) Interval estimation for the two-parameter Pareto distribution based on record values. *Selçuk J Appl Math* pp 149–161
6. Awad AM, Raqab MZ (2000) Prediction intervals for the future record values from exponential distribution: Comparative study. *J Stat Comput Simul* 65:325–340
7. Bayarri MJ, DeGroot MH (1988) Discussion: Auxiliary parameters and simple likelihood functions. In: Berger JO, Wolpert RL (eds) *The Likelihood Principle*. Inst. Math. Statistics. Hayward (Cal), p 160.3–160.7
8. Beirlant J, Teugels JL (1992) Modeling large claims in non-life insurance. *Insur Math Econ* 11(1):17–29

9. Beirlant J, Goegebeur Y, Segers J et al (2004) *Statistics of Extremes: Theory and Applications*. Wiley, Amsterdam
10. Berred AM (1998) Prediction of record values. *Comm Stat Theory Methods* 27(9):2221–2240
11. Bevere L, Weigel A (2021) sigma 1/2021 - Natural catastrophes in 2020. Swiss Re Management Ltd, Swiss Re Institute. <https://www.swissre.com/institute/research/sigma-research/sigma-2021-01.html>. Accessed 15 Nov 2023
12. Bladt M, Albrecher H, Beirlant J (2020) Combined tail estimation using censored data and expert information. *Scand Actuar J* 6:503–525
13. Brazauskas V, Kleefeld A (2009) Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insur Mat Econ* 45(3):424–435
14. Brazauskas V, Kleefeld A (2016) Modeling severity and measuring tail risk of Norwegian fire claims. *N Am Actuar J* 20(1):1–16
15. Castaño-Martínez A, López-Blázquez F, Pigueras G et al (2020) A method for constructing and interpreting some weighted premium principles. *ASTIN Bull* 50(3):1037–1064
16. Chandler KN (1952) The distribution and frequency of record values. *J R Stat Soc Ser B (Methodological)* 14(2):220–228
17. Dziubdziela W, Kopociński B (1976) Limiting properties of the  $k$ -th record values. *Applicationes Mathematicae* 15(2):187–190
18. Embrechts P, Klüppelberg C, Mikosch T (2012) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin
19. Empacher C, Kamps U, Volovskiy G (2023) Statistical prediction of future sports records based on record values. *Stats* 6:131–147
20. Feuerverger A, Hall P (1998) On statistical inference based on record values. *Extremes* 1:169–190
21. Kamps U (1995) *A Concept of Generalized Order Statistics*. Teubner, Stuttgart
22. Madi MT, Raqab MZ (2004) Bayesian prediction of temperature records using the Pareto model. *Environmetrics* 15:701–710
23. McNeil AJ (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bull* 27(1):117–137
24. Nagaraja HN (1988) Record values and related statistics - a review. *Comm Stat Theory Methods* 17:2223–2238
25. Nevzorov V (2001) *Records: Mathematical Theory*. American Mathematical Society, Providence
26. Pfaff B, Zivot E, McNeil A, et al (2018) Extreme values in R. R-package. <https://cran.r-project.org/web/packages/evir/evir.pdf>. Accessed 16 Nov 2023
27. Raqab MZ (2006) Nonparametric prediction intervals for the future rainfall records. *Environmetrics* 17(5):457–464
28. Raqab MZ, Ahmadi J, Doostparast M (2007) Statistical inference based on record data from Pareto model. *Statistics* 41:105–118
29. Raschke M (2020) Alternative modelling and inference methods for claim size distributions. *Ann Actuar Sci* 14(1):1–19
30. Resnick SI (1997) Discussion of the Danish data on large fire insurance losses. *ASTIN Bull* 27(1):139–151
31. Volovskiy G, Kamps U (2020) Maximum observed likelihood prediction of future record values. *TEST* 29(4):1072–1097
32. Volovskiy G, Kamps U (2020) Maximum product of spacings prediction of future record values. *Metrika* 83:853–868
33. Volovskiy G, Kamps U (2023) Comparison of likelihood-based predictors of future Pareto and Lomax record values in terms of Pitman closeness. *Comm Stat Theory Methods* 52(6):1905–1922