Maximum observed likelihood prediction of future record values

Grigoriy Volovskiy[1] · **Udo Kamps**[1]

Abstract

Point prediction of future upper record values is considered. For an underlying absolutely continuous distribution with strictly increasing cumulative distribution function, the general form of the predictor obtained by maximizing the observed predictive likelihood function is established. The results are illustrated for the exponential, extreme-value and power-function distributions, and the performance of the obtained predictors is compared to that of maximum likelihood predictors on the basis of the mean squared error and the Pitman's measure of closeness criteria. For exponential and extreme-value distributions, it is shown that under slight restrictions, the maximum observed likelihood predictor outperforms the maximum likelihood predictor in terms of both performance criteria.

Keywords Maximum likelihood prediction · Maximum observed likelihood prediction · Upper record values · Exponential distribution · Extreme-value distribution · Power-function distribution

Mathematics Subject Classification 62F99 · 62M20

1 Introduction

Let $X_1, X_2, \ldots$ be an infinite sequence of independent and identically distributed (i.i.d.) random variables with continuous cumulative distribution function (cdf) F. An observation X_j is called an *(upper) record value*, provided it is greater than all previously observed values. More specifically, defining the record times as

✉ Grigoriy Volovskiy
volovskiy@isw.rwth-aachen.de

Udo Kamps
kamps@isw.rwth-aachen.de

[1] Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany

Springer

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

the sequence $(R_n)_{n \in \mathbb{N}} = (X_{L(n)})_{n \in \mathbb{N}}$ is referred to as the sequence of (upper) record values based on $(X_n)_{n \in \mathbb{N}}$ (see Arnold et al. 1998; Nevzorov 2001).

Record values, first studied by Chandler (1952), provide a natural model for the sequence of successive extremes in an i.i.d. sequence of random variables. In mathematical reliability theory, record values appear in the context of minimal repair systems (see Gupta and Kirmani 1988). There is also a close connection between the occurrence times of a nonhomogeneous Poisson process (NHPP) and record values. Indeed, by the results in Gupta and Kirmani (1988), under very mild conditions, the epoch times of a NHPP and record values are equal in distribution.

The problem of predicting a future record value R_s based on the observed record values $R_1, \ldots, R_r, r < s$ has been studied by several authors. As far as non-Bayesian prediction is concerned, most of the proposed predictors have been derived by applying well-known prediction procedures that have previously been applied in the context of other models of ordered data. Specifically for one-sample prediction of record values, we refer to Raqab (2007), where the best linear unbiased predictor, the best linear equivariant predictor, the maximum likelihood predictor (MLP) as well as the conditional median predictor of the sth record value R_s based on a Type II left-censored sample from the two-parameter exponential distribution are derived. His results supplement and generalize the results of Ahsanullah (1980), Basak and Balakrishnan (2003) and Nagaraja (1986, Section 4). A comparative study of several predictors of the sth record value R_s based on the first r observed record values from the one-parameter exponential distribution can be found in Awad and Raqab (2000). Maximum likelihood prediction of future Pareto record values is studied in Raqab et al. (2007). Moreover, since the record value model is contained in the generalized order statistics model (see Kamps 1995), all results pertaining to prediction of future generalized order statistics can be specialized to solve the prediction problem for record values (see, e.g., Burkschat 2009). Bayesian prediction methods for future record values were first discussed by Dunsmore (1983) and have subsequently been applied to various distribution families. For these results, we refer to, e.g., Madi and Raqab (2004), Ahmadi and Doostparast (2006) and Nadar and Kızılaslan (2015).

As specifically for likelihood-based prediction methods, the *maximum likelihood prediction* procedure (see Kaminsky and Rhodin 1985) has received a great deal of attention in the literature. Applying a likelihood-based prediction method in the context of an ordered data model has so far been synonymous with applying the maximum likelihood prediction procedure. In this paper, an alternative maximum likelihood predictor, the *maximum observed likelihood predictor (MOLP)*, is proposed and subsequently applied to predict future record values. Contrary to the maximum likelihood prediction procedure, the new method allows to derive the general form of the predictor as a function of the estimator of the underlying distributional parameters. Moreover, the obtained predictors outperform the MLP, which is illustrated by means of comparing MOLPs and MLPs of future exponential and extreme-value record values in terms of mean squared error and Pitman's measure of closeness. For properties of the MOLP, when the underlying distribution is assumed to be a Pareto, Lomax or Weibull distribution, we refer to Volovskiy (2018).

2 Maximum observed likelihood prediction procedure

Let X, Y be absolutely continuous random variables with values in $\mathbb{R}^p$ and $\mathbb{R}$, respectively, and joint probability density function $f_{\boldsymbol{\theta}}^{X,Y}$ known up to a parameter vector $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$. Random variable X models observed data, while Y stands for a yet-not-observed value to be predicted using a predictor $\pi(X)$. In non-Bayesian prediction setups, a natural approach to finding a predictor for Y based on X has been to define a generalized (parametric) likelihood function that can be used to solve statistical problems involving both fixed unknown parameters and unobserved random variables. In Bayarri et al. (1987), the authors consider the functions

$$L_{\mathrm{rv}}(y, \boldsymbol{\theta}|\boldsymbol{x}) = f_{\boldsymbol{\theta}}^{X,Y}(\boldsymbol{x}, y) \quad \text{and} \quad L_{\mathrm{obs}}(y, \boldsymbol{\theta}|\boldsymbol{x}) = f_{\boldsymbol{\theta}}^{X|Y}(\boldsymbol{x}|y),$$

in what follows to be called *predictive likelihood function (PLF)* and *observed predictive likelihood function (OPLF)*, respectively, as possible extensions of the classical parametric likelihood function and implement the maximum likelihood principle to obtain an estimate for $\boldsymbol{\theta}$ and a prediction value for Y. They compare the proposed likelihood functions by comparing the estimates and prediction values obtained from them. By way of a slightly contrived example (see Bayarri et al. 1987, Section 2), the authors demonstrate that the maximum likelihood method applied to either L_{rv} or L_{obs} does not yield reasonable results in general, which led them to conclude that no general definition of a likelihood function can be given, only to argue in Bayarri and DeGroot (1988) in favor of L_{obs}.

There has also been an attempt to justify the use of either L_{rv} or L_{obs} for deriving predictive inferences by using arguments from the theoretical foundations of statistical inference. In parametric inference, Fisher's likelihood function is pivotal to the formulation of the likelihood principle, and it is Birnbaum's theorem (see Birnbaum 1962), which establishes the equivalence of the likelihood principle and the sufficiency and conditionality principles, that can be seen as providing the theoretical justification for the choice of Fisher's likelihood function as a basis for parametric statistical analysis. For an in-depth discussion of the likelihood principle and Birnbaum's theorem, we refer the reader to the monograph by Berger and Wolpert (1988). It has been recognized that Birnbaum's result can serve as a guidance in generalizing the parametric likelihood beyond the case of parametric inference by requiring that the likelihood function be specified in such a way that the equivalence of the likelihood principle and the suitably modified sufficiency and conditionality principles continues to hold. This program was realized by Bjørnstad (1996) and Nayak and Kundu (2002). However, while the analysis in Bjørnstad (1996) provides a justification for L_{rv} as a general specification of the likelihood function, the discussion in Nayak and Kundu (2002) favors L_{obs}. Evidently, the generalizations of the sufficiency and conditionality principles proposed by Bjørnstad (1996) and Nayak and Kundu (2002) do not accord.

Certainly, the example in Bayarri et al. (1987, Section 2) can be understood to advise caution against careless application of classical likelihood methods when the likelihood function is extended to include random variables. However, one may also take the view that likelihood functions L_{rv} and L_{obs} are tools, albeit not of universal

applicability, that can serve to derive predictors. This intention was behind the introduction of the maximum likelihood prediction procedure by Kaminsky and Rhodin (1985), which we briefly recall in the following definition.

Definition 2.1 Suppose $\pi_{\mathrm{MLP}} : (\mathbb{R}^p, \mathcal{B}^p) \to (\mathbb{R}, \mathcal{B})$ and $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} : (\mathbb{R}^p, \mathcal{B}^p) \to (\Theta, \mathcal{B}^d_{|\Theta})$ are functions such that for any $\boldsymbol{x} \in \mathbb{R}^p$

$$L_{\mathrm{rv}}(\pi_{\mathrm{MLP}}(\boldsymbol{x}), \hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{x})|\boldsymbol{x}) = \max_{(y,\boldsymbol{\theta})\in\mathbb{R}\times\Theta} L_{\mathrm{rv}}(y, \boldsymbol{\theta}|\boldsymbol{x}).$$

Then, we call $\pi_{\mathrm{MLP}}(\boldsymbol{X})$ and $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{X})$ *maximum likelihood predictor (MLP)* of Y and *predictive maximum likelihood estimator (PMLE)* of $\boldsymbol{\theta}$, respectively.

Since its introduction, the maximum likelihood prediction procedure has become a standard method in models of ordered data. It was applied to the prediction of record values (see, e.g., Basak and Balakrishnan 2003), prediction of failure times of censored units in a progressive censoring procedure (see, e.g., Balakrishnan and Cramer 2014, Chapter 16) and generalized order statistics (see, e.g., Raqab 2001). Further references will be provided in Sect. 3. Apart from prediction based on ordered data, the method was applied to solve prediction problems in actuarial mathematics (see Kaminsky 1987). Contrary to prediction based on L_{rv}, to the best of our knowledge, prediction based on maximization of L_{obs} has received no attention apart from the articles focused on the foundations of statistics cited above. We propose to reconsider the approach by introducing the following L_{obs}-based prediction procedure:

Definition 2.2 Suppose $\pi_{\mathrm{MOLP}} : (\mathbb{R}^p, \mathcal{B}^p) \to (\mathbb{R}, \mathcal{B})$ and $\hat{\boldsymbol{\theta}}_{\mathrm{MOL}} : (\mathbb{R}^p, \mathcal{B}^p) \to (\Theta, \mathcal{B}^d_{|\Theta})$ are functions such that for any $\boldsymbol{x} \in \mathbb{R}^p$,

$$L_{\mathrm{obs}}(\pi_{\mathrm{MOLP}}(\boldsymbol{x}), \hat{\boldsymbol{\theta}}_{\mathrm{MOL}}(\boldsymbol{x})|\boldsymbol{x}) = \max_{\substack{(y,\boldsymbol{\theta})\in\mathbb{R}\times\Theta: \\ f^Y_{\boldsymbol{\theta}}(y)>0}} L_{\mathrm{obs}}(y, \boldsymbol{\theta}|\boldsymbol{x}).$$

Then, $\pi_{\mathrm{MOLP}}(\boldsymbol{X})$ and $\hat{\boldsymbol{\theta}}_{\mathrm{MOL}}(\boldsymbol{X})$ are termed *maximum observed likelihood predictor (MOLP)* of Y and *predictive maximum observed likelihood estimator (PMOLE)* of $\boldsymbol{\theta}$, respectively.

The parameter $\boldsymbol{\theta}$ may (partly) disappear from the function L_{obs} (see an example in Bayarri et al. 1987, Section 2), in which case L_{obs} does not provide guidance as to the (complete) choice of an estimator for $\boldsymbol{\theta}$, i.e., $\hat{\boldsymbol{\theta}}_{\mathrm{MOL}}$ may, except for the restriction that it takes values in Θ, be (in part) arbitrary. Thus, apart from the situation when the set of parameters present in L_{obs} coincides with the subset of parameters that are of inferential interest, in general, the predictive maximum observed likelihood estimator of $\boldsymbol{\theta}$ aims at ensuring the predictor of Y exists uniquely. A similar conclusion is also valid with respect to the predictive maximum likelihood estimator, which stems from the fact that, in general, $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is determined by Y and thus hardly can be considered a "sound" estimator of $\boldsymbol{\theta}$. The maximum observed likelihood prediction procedure will be applied to the prediction problem of future record values in the following section.

3 Prediction of future record values

Let $(R_n)_{n=1}^\infty$ be the sequence of record values in a sequence of i.i.d. random variables with absolutely continuous cdf F_θ and density function f_θ, $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$. In the present section, we aim to provide sufficient conditions for the existence of the MOLP of R_s based on $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$, $r, s \in \mathbb{N}, r < s$.

It turns out that due to the structure of the observed predictive likelihood function, the problem of finding the MOLP and the PMOLE can be reduced to that of finding the PMOLE. In order to derive the observed predictive likelihood function, we will need explicit expressions for the density functions of the distributions of $\boldsymbol{R}_\star$ and R_s as well as of the conditional distribution of R_s given $R_r = x$. These are summarized in the following lemma (see, e.g., Arnold et al. 1998). In what follows, we use the notational convention that for an interval $I \subseteq \mathbb{R}$ and $n \in \mathbb{N}$, $I_<^n = \{(x_1, \ldots, x_n) \in I^n \mid x_1 < \cdots < x_n\}$. Moreover, the left and right endpoints of the support of a distribution with cdf F are denoted, respectively, by $\alpha(F)$ and $\omega(F)$. Throughout, for cdf F with density function f, h denotes the hazard rate function defined by $h(x) = f(x)/(1 - F(x))$, for $x < \omega(F)$.

Lemma 3.1 *Let $(R_n)_{n=1}^\infty$ be the sequence of record values in a sequence of i.i.d. random variables with absolutely continuous cdf F and density function f. Then, for $r, s \in \mathbb{N}$, $r < s$, the density functions of the distributions of $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$, R_s as well as the conditional distribution of R_s given $R_r = x$, $x \in (-\infty, \omega(F))$, are given by*

$$f^{\boldsymbol{R}_\star}(\boldsymbol{x}) = \left(\prod_{i=1}^{r-1} h(x_i) \right) f(x_r) \mathbb{1}_{[\alpha(F), \omega(F))^r}(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, \ldots, x_r) \in \mathbb{R}_<^r,$$

$$f^{R_s}(y) = \frac{1}{(s-1)!} f(y)(-\ln(1 - F(y)))^{s-1} \mathbb{1}_{[\alpha(F), \omega(F))}(y), \quad y \in \mathbb{R},$$

$$f^{R_s|R_r}(y|x) = \frac{f(y)/(s-r-1)!}{1 - F(x)} \left(\ln\left(\frac{1 - F(x)}{1 - F(y)} \right) \right)^{s-r-1} \mathbb{1}_{(x, \omega(F))}(y), \quad y \in \mathbb{R}.$$

Since, by assumption, for all $\theta \in \Theta$, the underlying cdf F_θ is continuous, the sequence $(R_n)_{n=1}^\infty$ possesses the Markov property. Hence, by adopting the convention that $0/0 := 0$, the observed predictive likelihood function of R_s and θ given $\boldsymbol{R}_\star = \boldsymbol{x}_\star$, $\boldsymbol{x}_\star = (x_1, \ldots, x_r) \in \mathbb{R}_<^r$, can be expressed as

$$L_{\text{obs}}(x_s, \theta | \boldsymbol{x}_\star) = f_\theta^{\boldsymbol{R}_\star}(\boldsymbol{x}_\star) f_\theta^{R_s|R_r}(x_s|x_r)/f_\theta^{R_s}(x_s), \quad x_s \in \mathbb{R}, \theta \in \Theta.$$

From this expression along with Lemma 3.1, we obtain that for a given $\boldsymbol{x}_\star \in \mathbb{R}_<^r$, and $x_s \in \mathbb{R}$, $\theta \in \Theta$, the observed predictive likelihood function satisfies

$$L_{\text{obs}}(x_s, \theta | \boldsymbol{x}_\star) \propto \left(\prod_{i=1}^r \frac{h_\theta(x_i)}{(-\ln(1 - F_\theta(x_r)))} \right) (1 - G_{\theta, x_s}(x_r))^{s-r-1} (G_{\theta, x_s}(x_r))^r$$

$$\times \mathbb{1}_{[\alpha(F_\theta), \omega(F_\theta))^r \times (x_r, \omega(F_\theta))}(\boldsymbol{x}_\star, x_s), \tag{1}$$

where for $x \le y < \omega(F_\theta)$, $G_{\theta, y}(x) = \ln(1 - F_\theta(x))/\ln(1 - F_\theta(y))$.

Remark 3.1 (i) The OPLF can be rewritten as

$$L_{\text{obs}}(x_s, \boldsymbol{\theta} | \boldsymbol{x}_\star) \propto \prod_{i=1}^{r} G'_{\boldsymbol{\theta}, x_s}(x)|_{x=x_i} \left(1 - G_{\boldsymbol{\theta}, x_s}(x_r)\right)^{s-r-1}$$

$$\times \mathbb{1}_{[\alpha(F_{\boldsymbol{\theta}}), \omega(F_{\boldsymbol{\theta}}))^r \times (x_r, \omega(F_{\boldsymbol{\theta}}))}(\boldsymbol{x}_\star, x_s), \quad x_s \in \mathbb{R}, \ \boldsymbol{\theta} \in \Theta. \quad (2)$$

This representation is related to the fact that the conditional distribution of $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$ given $R_s = y$, $y \in (\alpha(F_{\boldsymbol{\theta}}), \omega(F_{\boldsymbol{\theta}}))$, coincides with the distribution of the first r ordinary order statistics from a sample of $s-1$ i.i.d. random variables with cdf $G_{\boldsymbol{\theta}, y}$ (see also Keseling 1999, Remark 1.17). From representation (2), we see that a sub-parameter θ_i of the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$ is not estimable by the method of observed predictive likelihood maximization if it does not appear in any of the functions $(x, y) \mapsto G_{\boldsymbol{\theta}, y}(x)$, $(x, y) \in (\alpha(F_{\boldsymbol{\theta}}), \omega(F_{\boldsymbol{\theta}}))^2_<$ parameterized by $\boldsymbol{\theta} \in \Theta$. As an example, consider the two-parameter exponential distribution. Then, $\boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$ and, for $y > \mu$, $G_{\boldsymbol{\theta}, y}(x) = \frac{x-\mu}{y-\mu}$, for $x \leq y$, and $G_{\boldsymbol{\theta}, y}(x) = 1$ otherwise. Consequently, the method of observed predictive likelihood maximization cannot produce a meaningful estimator for the sub-parameter σ. However, this parameter dropout does not affect the usefulness of the method as a vehicle for deriving predictors for future record values as evidenced by Theorem 3.1.

(ii) For $k \in \mathbb{N}$, the observed predictive likelihood function (1) coincides with the observed predictive likelihood function of $R_s^{(k)}$ and $\boldsymbol{\theta}$ given $(R_1^{(k)}, \ldots, R_r^{(k)}) = (x_1, \ldots, x_r)$, where $(R_n^{(k)})_{n \in \mathbb{N}}$ denotes the sequence of kth record values in a sequence of i.i.d. random variables with cdf $F_{\boldsymbol{\theta}}$ (see Dziubdziela and Kopociński 1976). This follows from the fact that kth record values in a sequence of i.i.d. random variables with cdf F are equal in distribution to record values in a sequence of i.i.d. random variables with cdf $F_{1:k} = 1 - (1 - F)^k$ (see Arnold et al. 1998, p. 43).

In the following, for $\boldsymbol{x}_\star \in \mathbb{R}^r_<$, $\Psi(\cdot | \boldsymbol{x}_\star)$ denotes the function given by

$$\Psi(\boldsymbol{\theta} | \boldsymbol{x}_\star) = \left(\prod_{i=1}^{r} \frac{h_{\boldsymbol{\theta}}(x_i)}{(-\ln(1 - F_{\boldsymbol{\theta}}(x_r)))} \right) \mathbb{1}_{[\alpha(F_{\boldsymbol{\theta}}), \omega(F_{\boldsymbol{\theta}}))^r}(\boldsymbol{x}_\star), \quad \boldsymbol{\theta} \in \Theta. \quad (3)$$

In Theorem 3.1, the assumption $r + 1 < s$ is made. This is due to the fact that, for $s = r + 1$, if a MOLP exists, it is necessarily given by $\pi_{\text{MOLP}}^{(s)} = R_r$. Thus, in this case, the maximum observed likelihood prediction method does not produce a reasonable predictor. We refer to Remark 3.2 (iii) for more details.

Theorem 3.1 *For $s \geq 3$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. random variables with cdf $F_{\boldsymbol{\theta}}$, which, for all $\boldsymbol{\theta} \in \Theta$, is assumed to be absolutely continuous and strictly increasing on its support. Moreover, for $r \in \mathbb{N}$, $r < s - 1$, let $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$ and let $\hat{\boldsymbol{\theta}} : (\mathbb{R}^r, \mathcal{B}^r) \to (\Theta, \mathcal{B}^d_{|\Theta})$ be a function with the property*

$$\Psi(\hat{\boldsymbol{\theta}}(\boldsymbol{x}_\star) | \boldsymbol{x}_\star) = \max_{\boldsymbol{\theta} \in \Theta} \Psi(\boldsymbol{\theta} | \boldsymbol{x}_\star), \quad \boldsymbol{x}_\star \in \mathbb{R}^r_<, \quad (4)$$

where Ψ is given in (3). *Then, a MOLP* $\pi_{\mathrm{MOLP}}^{(s)}$ *of R_s and a PMOLE* $\hat{\boldsymbol{\theta}}_{\mathrm{MOL}}$ *of $\boldsymbol{\theta}$ based on $\boldsymbol{R}_\star$ are given by*

$$\pi_{\mathrm{MOLP}}^{(s)} = F_{\hat{\boldsymbol{\theta}}_{\mathrm{MOL}}}^{-1} \left(1 - (1 - F_{\hat{\boldsymbol{\theta}}_{\mathrm{MOL}}}(R_r))^{\frac{s-1}{r}}\right) \quad and \quad \hat{\boldsymbol{\theta}}_{\mathrm{MOL}} = \hat{\boldsymbol{\theta}}(\boldsymbol{R}_\star). \qquad (5)$$

Proof The aim is to maximize the observed predictive likelihood function of R_s and $\boldsymbol{\theta}$ given $\boldsymbol{R}_\star = \boldsymbol{x}_\star$, $\boldsymbol{x}_\star \in \mathbb{R}_<^r$. Fix some $\boldsymbol{x}_\star \in \mathbb{R}_<^r$. Then, by the assumed property (4) of $\hat{\boldsymbol{\theta}}$, we have that, for $\boldsymbol{\theta} \in \Theta$ and $x_s \in (x_r, \omega(F_{\boldsymbol{\theta}}))$,

$$\begin{aligned} L_{\mathrm{obs}}(x_s, \boldsymbol{\theta}|\boldsymbol{x}_\star) &\propto \Psi(\boldsymbol{\theta}|\boldsymbol{x}_\star) \left(1 - G_{\boldsymbol{\theta}, x_s}(x_r)\right)^{s-r-1} \left(G_{\boldsymbol{\theta}, x_s}(x_r)\right)^r \\ &\leq \Psi(\hat{\boldsymbol{\theta}}(\boldsymbol{x}_\star)|\boldsymbol{x}_\star) \left(1 - G_{\boldsymbol{\theta}, x_s}(x_r)\right)^{s-r-1} \left(G_{\boldsymbol{\theta}, x_s}(x_r)\right)^r. \end{aligned}$$

Now, using the well-known expression for the mode of the probability density of a beta distribution with parameters $s - r$ and $r + 1$, which by assumption are larger than 1, as well as the assumed strict monotonicity of $F_{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta} \in \Theta$, we obtain that, for any $\boldsymbol{\theta} \in \Theta$, the function

$$l_{\boldsymbol{\theta}}(x_s) = \left(1 - G_{\boldsymbol{\theta}, x_s}(x_r)\right)^{s-r-1} \left(G_{\boldsymbol{\theta}, x_s}(x_r)\right)^r,$$

$x_s \in (x_r, \omega(F_{\boldsymbol{\theta}}))$, possesses a unique maximum point, which is obtained as a unique solution of the equation $G_{\boldsymbol{\theta}, x_s}(x_r) = \frac{r}{s-1}$ with respect to $x_s \in (x_r, \omega(F_{\boldsymbol{\theta}}))$. The solution of this equation is given by $x_s(\boldsymbol{\theta}, \boldsymbol{x}_\star) = F_{\boldsymbol{\theta}}^{-1}(1 - (1 - F_{\boldsymbol{\theta}}(x_r))^{\frac{s-1}{r}})$. Moreover, we have that

$$l_{\boldsymbol{\theta}}(x_s(\boldsymbol{\theta}, \boldsymbol{x}_\star)) = ((s - r - 1)/(s - 1))^{s-r-1}(r/(s - 1))^r$$

independently of $\boldsymbol{\theta}$. Combining the preceding results yields that $L_{\mathrm{obs}}(x_s, \boldsymbol{\theta}|\boldsymbol{x}_\star) \leq L_{\mathrm{obs}}(x_s(\hat{\boldsymbol{\theta}}(\boldsymbol{x}_\star), \boldsymbol{x}_\star), \hat{\boldsymbol{\theta}}(\boldsymbol{x}_\star)|\boldsymbol{x}_\star)$, for all $x_s \in (x_r, \omega(F_{\boldsymbol{\theta}}))$, $\boldsymbol{\theta} \in \Theta$. Since $\boldsymbol{x}_\star$ was arbitrary, this shows that the predictor of R_s and the estimator of $\boldsymbol{\theta}$ as defined in (5) are indeed the MOLP of R_s and the PMOLE of $\boldsymbol{\theta}$. Thus, the proof is complete. $\qquad \square$

Remark 3.2 (i) Since the quantity $l_{\boldsymbol{\theta}}(x_s(\boldsymbol{\theta}, \boldsymbol{x}_\star))$ in the proof of Theorem 3.1 does not depend on $\boldsymbol{\theta}$, the existence of a function $\hat{\boldsymbol{\theta}}$ satisfying (4) is also necessary for the existence of a MOLP.

(ii) On inspecting the proof of Theorem 3.1, we find that under the stated assumptions, if any two measurable functions $\hat{\boldsymbol{\theta}}_i : (\mathbb{R}^r, \mathcal{B}^r) \to (\Theta, \mathcal{B}_{|\Theta}^d)$, $i = 1, 2$, satisfying (4) with $\hat{\boldsymbol{\theta}}$ replaced by $\hat{\boldsymbol{\theta}}_i$, $i = 1, 2$, coincide $P_{\boldsymbol{\theta}}^{\boldsymbol{R}_\star}$-almost surely, then the MOLP of R_s based on $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$ is also $P_{\boldsymbol{\theta}}^{\boldsymbol{R}_\star}$-almost surely uniquely determined.

(iii) It is possible to obtain a predictor by the method of observed predictive likelihood maximization also in the case that $s = r + 1$ by replacing the factor $\mathbb{1}_{(x_r, \omega(F_{\boldsymbol{\theta}}))}(x_s)$ in (1) by $\mathbb{1}_{[x_r, \omega(F_{\boldsymbol{\theta}}))}(x_s)$. However, this modification leads to the MOLP $\pi_{\mathrm{MOLP}}^{(s)} = R_r$, i.e., the last observed records value, which is certainly not a useful predictor. Aiming at comparing the MOLP with the MLP, we point out that, in all the

examples we will consider in the following subsections, the predictor produced by the method of predictive likelihood maximization is also given by the last observed record value. This problem can be overcome by using another method of prediction or by applying appropriate prediction intervals (cf., e.g., Awad and Raqab 2000).

(iv) Since function Ψ in (3) does not depend on s, the PMOLE of θ does not depend on s, either. Hence, $\hat{\theta}_{\text{MOL}}$ does not depend on which future record value one aims at predicting. We point out that the PMLE is not guaranteed to be free of the deficiency of depending on which future record is to be predicted. The occurrence of this shortcoming with estimators produced by the method of predictive likelihood maximization was observed by Bayarri et al. (1987, p. 8).

In the following subsections, we derive the MOLP and the MLP for future record values from different underlying distributions and compare their performance in terms of the mean squared error as well as Pitman's measure of closeness. As far as prediction in models of ordered data is concerned, Nagaraja (1986, Sections 3 and 4) was one of the first to use Pitman's measure of closeness as an alternative criterion to the mean squared error in a comparative study of predictors of future order statistics and record values.

3.1 Exponential distribution

Here, we assume that $(R_n)_{n \in \mathbb{N}}$ is the sequence of record values in a sequence of i.i.d. two-parameter exponential random variables. The density, cumulative distribution and quantile functions of the exponential distribution $\mathcal{E}xp(\mu, \sigma)$ with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma \in \mathbb{R}_+$ are given by $f_{\boldsymbol{\theta}}(x) = \exp\{-(x - \mu)/\sigma\}/\sigma$, $F_{\boldsymbol{\theta}}(x) = 1 - \exp\{-(x - \mu)/\sigma\}$, $x \in [\mu, \infty)$ and $F_{\boldsymbol{\theta}}^{-1}(x) = \mu - \sigma \ln(1-x)$, $x \in [0, 1)$, where $\boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$. Next, for $r, s \in \mathbb{N}, r < s-1$, we derive the MOLP and present the MLP of R_s based on $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$.

The results concerning the form of the MOLP of R_s are contained in the following proposition.

Proposition 3.1 *For $s \in \mathbb{N}$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. two-parameter exponential random variables.*

(i) *If μ is known, for $r, s \in \mathbb{N}$, $1 \le r < s - 1$, the unique MOLP of R_s based on $\boldsymbol{R}_\star$ is given by*

$$\pi_{\text{MOLP}}^{(s)} = R_r + (R_r - \mu)\frac{s - r - 1}{r}.$$

(ii) *If μ is unknown, for $r, s \in \mathbb{N}$, $2 \le r < s - 1$, the unique MOLP of R_s based on $\boldsymbol{R}_\star$ is given by*

$$\pi_{\text{MOLP}}^{(s)} = R_r + (R_r - R_1)\frac{s - r - 1}{r}.$$

The PMOLE of μ has the form $\hat{\mu}_{\text{MOL}} = R_1$.

Proof Assume that μ is unknown. With the above choice of $f_{\boldsymbol{\theta}}$ and $F_{\boldsymbol{\theta}}$, the function $\Psi(\cdot|\boldsymbol{x}_\star)$, $\boldsymbol{x}_\star \in \mathbb{R}^r_<$, in (3) becomes

$$\Psi(\boldsymbol{\theta}|\boldsymbol{x}_\star) = \frac{1}{(x_r - \mu)^r} \mathbb{1}_{(-\infty, x_1]}(\mu), \quad \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}_+. \tag{6}$$

As the scale parameter σ is not present in Ψ, we only need to find a maximizing function with respect to μ. Let $\hat{\boldsymbol{\theta}}(\boldsymbol{x}_\star) = (\hat{\sigma}(\boldsymbol{x}_\star), x_1)$, where $\hat{\sigma}$ is an arbitrary measurable function on $\mathbb{R}^r_<$ with values in $\mathbb{R}_+$. Then, assuming that $r \geq 2$, $\hat{\boldsymbol{\theta}}$ satisfies (4) with $\Psi(\cdot|\boldsymbol{x}_\star)$ given by (6). Combining this with the fact that

$$F_{\boldsymbol{\theta}}^{-1}\left(1 - (1 - F_{\boldsymbol{\theta}}(R_r))^{\frac{s-1}{r}}\right) = \mu + (R_r - \mu)\frac{s-1}{r}$$

yields that

$$\pi_{\text{MOLP}}^{(s)} = R_1 + (R_r - R_1)\frac{s-1}{r} = R_r + (R_r - R_1)\frac{s-r-1}{r}$$

is the unique maximum observed likelihood predictor of R_s based on $\boldsymbol{R}_\star$. Finally, we note that for known μ, the derivation of the predictor proceeds along the same lines. The details are omitted. □

From the results of Basak and Balakrishnan (2003), it follows that the MLP of R_s based on $\boldsymbol{R}_\star$ has the slightly different form

$$\pi_{\text{MLP}}^{(s)} = R_r + (R_r - \mu)\frac{s-r-1}{r+1},$$

if μ is known, and

$$\pi_{\text{MLP}}^{(s)} = R_r + (R_r - R_1)\frac{s-r-1}{r+1},$$

if μ is unknown.

3.1.1 Comparison based on the MSE

The mean squared errors of the MOLPs are given in the following lemma, where $\text{MSE}(\pi_{\text{MOLP}}^{(s)}) = E((\pi_{\text{MOLP}}^{(s)} - R_s)^2)$.

Lemma 3.2 *For $s \in \mathbb{N}$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. two-parameter exponential random variables.*

(i) *If μ is known, for $r, s \in \mathbb{N}$, $1 \leq r < s - 1$, the MSE of the MOLP of R_s based on $\boldsymbol{R}_\star$ is given by*

$$\text{MSE}(\pi_{\text{MOLP}}^{(s)}) = \sigma^2 \frac{(s-1)(s-r-1) + 2r}{r}.$$

(ii) *If μ is unknown, for $r, s \in \mathbb{N}$, $2 \leq r < s - 1$, the MSE of the MOLP of R_s based on $\boldsymbol{R_\star}$ is given by*

$$\text{MSE}(\pi_{\text{MOLP}}^{(s)}) = \sigma^2 \frac{(s - r + 1)(s - 1)}{r}.$$

Proof We present the derivation of the MSE in the case of unknown μ only. The proof of the other case proceeds along the same lines. We have

$$\text{MSE}(\pi_{\text{MOLP}}^{(s)})/\sigma^2 = \frac{1}{\sigma^2} E((R_s - R_r)^2) - 2 \frac{s - r - 1}{\sigma^2 r} E(R_s - R_r) E(R_r - R_1)$$

$$+ \frac{1}{\sigma^2} \left(\frac{s - r - 1}{r} \right)^2 E((R_r - R_1)^2)$$

$$= \frac{(s - r + 1)(s - 1)}{r}.$$

$\square$

By the results in Basak and Balakrishnan (2003), the MSE of the MLP is given by

$$\text{MSE}(\pi_{\text{MLP}}^{(s)}) = \sigma^2 \frac{s(s - r + 1)}{r + 1},$$

if μ is known, and by

$$\text{MSE}(\pi_{\text{MLP}}^{(s)}) = \sigma^2 \frac{s(3(s - r) + (s - r + 1)r - 1)}{(r + 1)^2},$$

if μ is unknown. As for the relative performance of the predictors in terms of mean squared error, we have the following result. For an unknown location parameter, the MOLP turns out to have smaller MSE than the MLP, throughout.

Proposition 3.2 *For $s \in \mathbb{N}$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. two-parameter exponential random variables. Moreover, let $\pi_{\text{MOLP}}^{(s)}$ and $\pi_{\text{MLP}}^{(s)}$ be the MOLP and the MLP of R_s based on $\boldsymbol{R_\star}$, respectively.*

(i) *If μ is known, for $r, s \in \mathbb{N}$, $1 \leq r < s - 1$,*

$$\text{MSE}(\pi_{\text{MOLP}}^{(s)}) < \text{MSE}(\pi_{\text{MLP}}^{(s)}) \quad \text{if and only if} \quad s < 3r + 1.$$

(ii) *If μ is unknown, for $r, s \in \mathbb{N}$, $2 \leq r < s - 1$,*

$$\text{MSE}(\pi_{\text{MOLP}}^{(s)}) < \text{MSE}(\pi_{\text{MLP}}^{(s)}).$$

Proof Again, we present only the proof of the case of unknown μ. Simple algebra yields

$$\frac{\text{MSE}(\pi_{\text{MLP}}^{(s)}) - \text{MSE}(\pi_{\text{MOLP}}^{(s)})}{\sigma^2} = \frac{(r - 1)(s + r + 1)(s - r - 1)}{r(r + 1)^2} > 0.$$
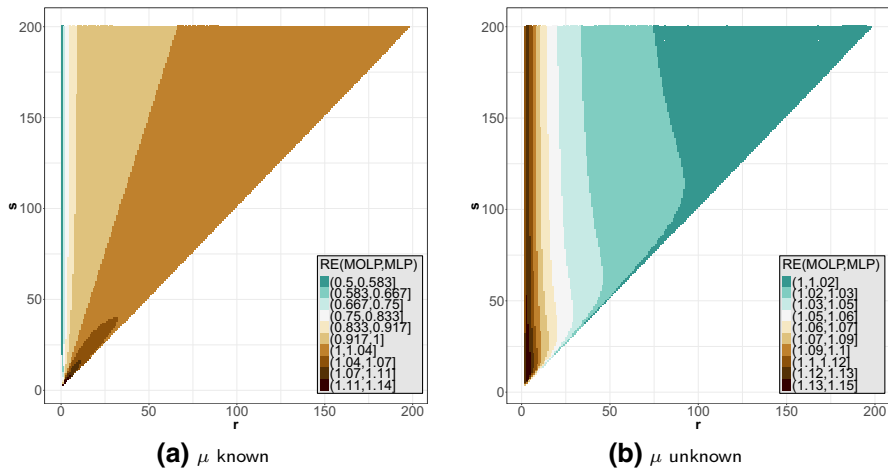
Fig. 1 Contour plot of the relative efficiency RE(MOLP, MLP) = $\text{MSE}(\pi_{\text{MLP}}^{(s)})/\text{MSE}(\pi_{\text{MOLP}}^{(s)})$ of the MLP of R_s relative to the MOLP of R_s based on two-parameter exponential record values $\boldsymbol{R}_\star$ for all valid combinations of r and s in the range 1–200 with $s = r + 1$ omitted

Hence, the MOLP has a smaller MSE than the MLP with no restrictions on r and s. Thus, the assertion is proved. □

As it can directly be seen in the above proof, the difference in Theorem 3.2(ii) increases in s. Figure 1 contains the contour plots of the relative efficiency

$$\text{RE(MOLP, MLP)} = \frac{\text{MSE}(\pi_{\text{MLP}}^{(s)})}{\text{MSE}(\pi_{\text{MOLP}}^{(s)})}$$

of $\pi_{\text{MLP}}^{(s)}$ relative to $\pi_{\text{MOLP}}^{(s)}$ based on $\boldsymbol{R}_\star$ for all valid combinations of r and s in the range of 1–200. Table 1 contains relative efficiencies of the predictors for selected values of r and s. It can be seen from the contour plots as well as the table that the relative efficiency of the MLP relative to the MOLP is the highest for small values of r and decreases as r increases. The gains in efficiency are in the high single-digit and low double-digit percentage range. From the perspective of practical applications, where rather small numbers of observed record values are common, this fact makes the MOLP an attractive alternative to the MLP.

3.1.2 Comparison based on Pitman's measure of closeness

In preparation for the comparison of the predictors in terms of Pitman's measure of closeness, we first compute the Pitman efficiency

$$\text{PE(MOLP, MLP)} = P\left(|R_s - \pi_{\text{MOLP}}^{(s)}| < |R_s - \pi_{\text{MLP}}^{(s)}|\right).$$

Table 1 Relative efficiencies RE(MOLP, MLP) = MSE($\pi_{\text{MLP}}^{(s)}$)/MSE($\pi_{\text{MOLP}}^{(s)}$) of the MLP of R_s relative to the MOLP of R_s based on two-parameter exponential record values $\boldsymbol{R_\star}$ for selected r and s

r	s						
---	$r+2$	$r+3$	$r+4$	$r+5$	$r+10$	$r+15$	$r+20$
(a) μ known							
2	1.143	1.111	1.053	1.000	0.854	0.795	0.764
5	1.094	1.111	1.103	1.087	1.011	0.966	0.939
10	1.056	1.074	1.079	1.077	1.047	1.021	1.003
15	1.039	1.055	1.060	1.061	1.048	1.032	1.019
20	1.030	1.043	1.048	1.050	1.044	1.034	1.024
25	1.025	1.036	1.040	1.043	1.040	1.033	1.026
50	1.013	1.019	1.022	1.024	1.025	1.024	1.021
80	1.008	1.012	1.014	1.016	1.017	1.017	1.016
(b) μ unknown							
2	1.086	1.111	1.120	1.123	1.124	1.122	1.120
5	1.080	1.111	1.125	1.132	1.136	1.133	1.130
10	1.052	1.074	1.086	1.092	1.099	1.098	1.095
15	1.038	1.055	1.064	1.069	1.076	1.076	1.074
20	1.029	1.043	1.051	1.055	1.062	1.062	1.061
25	1.024	1.036	1.042	1.046	1.052	1.053	1.052
50	1.013	1.019	1.022	1.025	1.029	1.030	1.030
80	1.008	1.012	1.014	1.016	1.019	1.020	1.020

The proof of the result is along the lines of that of the corresponding result in the comparison of the BLUP (best linear unbiased predictor) and the BLEP (best linear equivariant predictor) of order statistics from two-parameter exponential distributions presented in Nagaraja ([1986](), p. 14). In the following, $\mathcal{F}(m, n)$ denotes the F-distribution with parameters $m, n \in \mathbb{N}$. The corresponding cdf will be denoted by $F(\cdot | m, n)$. We also set $q = \frac{2(r+1)(s-r)}{(2r+1)(s-r-1)}$.

Lemma 3.3 *Let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. two-parameter exponential random variables. Moreover, let $\pi_{\text{MOLP}}^{(s)}$ and $\pi_{\text{MLP}}^{(s)}$ be the MOLP and the MLP of R_s based on $\boldsymbol{R_\star}$, respectively.*

(i) *If μ is known, for $r, s \in \mathbb{N}$, $1 \leq r < s - 1$, we have that*

$$P(|R_s - \pi_{\text{MOLP}}^{(s)}| < |R_s - \pi_{\text{MLP}}^{(s)}|) = F\left(q \, \Big| \, 2r, 2(s-r)\right).$$

(ii) *If μ is unknown, for $r, s \in \mathbb{N}$, $2 \leq r < s - 1$, we have that*

$$P(|R_s - \pi_{\text{MOLP}}^{(s)}| < |R_s - \pi_{\text{MLP}}^{(s)}|) = F\left(q \frac{r}{r-1} \, \Big| \, 2(r-1), 2(s-r)\right).$$

The following result characterizes the performance of the MLP relative to the MOLP based on Pitman's measure of closeness:

Proposition 3.3 *For $s \in \mathbb{N}$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. two-parameter exponential random variables. The MOLP of R_s based on $\boldsymbol{R}_\star$ outperforms the MLP of R_s based on $\boldsymbol{R}_\star$ in terms of Pitman's measure of closeness irrespectively of whether μ is known or not, i.e., for any r, s, satisfying $1 \le r \le s-2$ (μ known) or $2 \le r \le s-2$ (μ unknown), we have that*

$$P\left(|R_s - \pi^{(s)}_{\text{MOLP}}| < |R_s - \pi^{(s)}_{\text{MLP}}|\right) > P\left(|R_s - \pi^{(s)}_{\text{MOLP}}| > |R_s - \pi^{(s)}_{\text{MLP}}|\right).$$

Proof First, assume that μ is unknown. Let $X \sim \mathcal{F}(2(r-1), 2(s-r))$, $r \ge 2$, $r \le s-2$. Assume first that $r > 2$. Since for this choice of r and s the F-distribution satisfies the mode–median–mean inequality (see Groeneveld and Meeden 1977), it suffices to show that

$$E(X) < \frac{2r(r+1)(s-r)}{(2r+1)(r-1)(s-r-1)}. \tag{7}$$

Since $s - r \ge 2$, the expectation of X is finite, and we have that

$$E(X) = \frac{s-r}{s-r-1} < \frac{2r(r+1)(s-r)}{(2r+1)(r-1)(s-r-1)} \quad \text{if and only if} \quad 0 < 3r+1.$$

This establishes inequality (7). If $r = 2$, there is a closed-form expression for the median of X, which is given by $\text{med}(X) = (s-2)(2^{\frac{1}{s-2}} - 1)$. Since $\text{med}(X) \le E(X)$ if and only if $2 \le (1 + 1/(s-3))^{s-2}$, and the second inequality is satisfied for all $s \ge 4$, the validity of (7) yields the assertion. Next, assume that μ is known and let $X \sim \mathcal{F}(2r, 2(s-r))$. Again, reasoning as above, the aim is to show that $E(X) < \frac{2r(r+1)(s-r)}{(2r+1)r(s-r-1)}$. Now,

$$E(X) = \frac{s-r}{s-r-1} < \frac{2r(r+1)(s-r)}{(2r+1)r(s-r-1)} \quad \text{if and only if} \quad r > 0.$$

This yields the assertion for the case that $r > 1$. If $r = 1$, an analogous argument as in the first part of the proof yields the desired conclusion. $\square$

Figure 2 contains the contour plots of the Pitman efficiency of $\pi^{(s)}_{\text{MLP}}$ relative to $\pi^{(s)}_{\text{MOLP}}$ based on $\boldsymbol{R}_\star$ for all valid combinations of r and s in the range of 1–200. Table 2 contains the Pitman efficiencies of the predictors for selected values of r and s. In the parameter range examined, the Pitman efficiencies do not fall below 0.6 for values of r smaller than 15 and achieve maximum values of approximately 0.8 (known μ) and 0.9 (unknown μ). Again, from the perspective of practical applications, where rather small numbers of observed record values are common, this fact makes the MOLP an attractive alternative to the MLP.
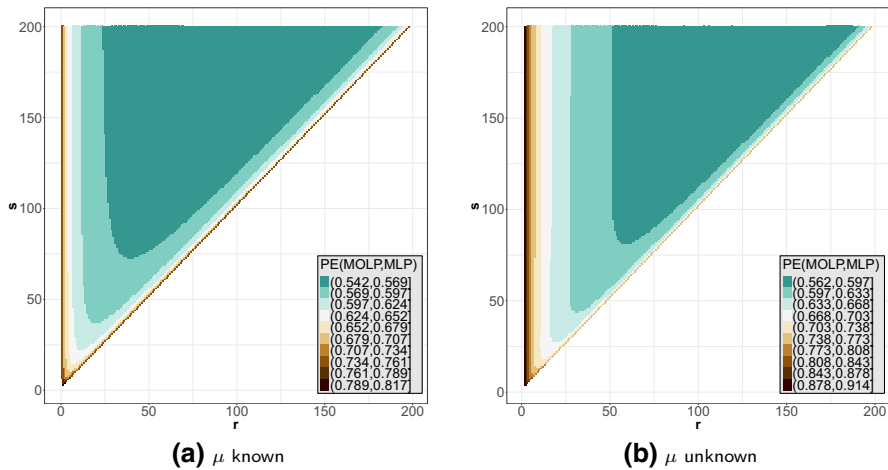
Fig. 2 Contour plot of the Pitman efficiency PE(MOLP, MLP) $= P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|)$ of the MLP of R_s relative to the MOLP of R_s based on two-parameter exponential record values $\boldsymbol{R}_\star$ for all valid combinations of r and s in the range of 1–200 with $s = r + 1$ omitted

Table 2 Pitman efficiencies PE(MOLP, MLP) $= P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|)$ of the MLP of R_s relative to the MOLP of R_s based on two-parameter exponential record values $R_1, \ldots, R_r$ for selected r and s

r	s						
	$r + 2$	$r + 3$	$r + 4$	$r + 5$	$r + 10$	$r + 15$	$r + 20$
(a) μ known							
2	0.791	0.752	0.735	0.726	0.708	0.702	0.700
5	0.765	0.718	0.696	0.683	0.659	0.651	0.647
10	0.752	0.700	0.675	0.661	0.631	0.621	0.616
15	0.747	0.693	0.667	0.651	0.619	0.608	0.602
20	0.744	0.689	0.663	0.646	0.612	0.600	0.594
25	0.743	0.687	0.660	0.643	0.608	0.595	0.588
50	0.739	0.682	0.654	0.636	0.598	0.584	0.576
80	0.738	0.680	0.651	0.634	0.594	0.579	0.571
(b) μ unknown							
2	0.913	0.906	0.905	0.905	0.906	0.907	0.907
5	0.826	0.801	0.792	0.789	0.786	0.787	0.788
10	0.785	0.747	0.731	0.723	0.713	0.712	0.712
15	0.770	0.726	0.707	0.696	0.679	0.676	0.676
20	0.762	0.714	0.693	0.681	0.660	0.655	0.654
25	0.757	0.707	0.685	0.672	0.648	0.642	0.639
50	0.747	0.692	0.667	0.651	0.620	0.610	0.605
80	0.743	0.687	0.660	0.643	0.608	0.596	0.590

3.2 Extreme-value distribution

Next, we assume that $(R_n)_{n\in\mathbb{N}}$ is the sequence of record values in a sequence of i.i.d. extreme-value random variables. The density, cumulative distribution and quantile functions of the extreme-value or reversed Gumbel distribution $EV(\mu, \sigma)$ with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma \in \mathbb{R}_+$ are given by $f_{\boldsymbol{\theta}}(x) = \exp\{(x-\mu)/\sigma - \exp\{(x-\mu)/\sigma\}\}/\sigma$, $F_{\boldsymbol{\theta}}(x) = 1 - \exp\{-\exp\{(x-\mu)/\sigma\}\}$, $x \in \mathbb{R}$, and $F_{\boldsymbol{\theta}}^{-1}(x) = \mu + \sigma\ln(-\ln(1-x))$, $x \in (0, 1)$, where $\boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$. Next, for $r, s \in \mathbb{N}$, $r < s - 1$, we derive the MOLP and the MLP of R_s based on $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$. As for the form of the MOLP of R_s based on $\boldsymbol{R}_\star$, we have the following result.

Proposition 3.4 *Let $R_1, \ldots, R_s$ be the first s, $s \geq 3$, record values in a sequence of i.i.d. extreme-value random variables. For $r \in \mathbb{N}$, $2 \leq r < s - 1$, the unique MOLP of R_s and the PMOLE of σ based on $\boldsymbol{R}_\star$ are given by*

$$\pi_{\text{MOLP}}^{(s)} = R_r + \hat{\sigma}_{\text{MOL}} \ln\left(\frac{s-1}{r}\right) \quad \text{and} \quad \hat{\sigma}_{\text{MOL}} = \frac{1}{r}\sum_{i=1}^{r-1}(R_r - R_i).$$

Proof With the above choice of $f_{\boldsymbol{\theta}}$ and $F_{\boldsymbol{\theta}}$, the function $\Psi(\cdot|\boldsymbol{x}_\star)$, $\boldsymbol{x}_\star \in \mathbb{R}_<^r$, in (3) becomes

$$\Psi(\boldsymbol{\theta}|\boldsymbol{x}_\star) = \frac{1}{\sigma^r}\exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{r-1}(x_r - x_i)\right\}, \quad \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}_+. \tag{8}$$

Since the location parameter μ is not present in Ψ, we only need to find a maximizing function with respect to σ. Let $\hat{\boldsymbol{\theta}}(\boldsymbol{x}_\star) = (\hat{\mu}(\boldsymbol{x}_\star), \frac{1}{r}\sum_{i=1}^{r-1}(x_r - x_i))$, where $\hat{\mu}$ is an arbitrary measurable function on $\mathbb{R}_<^r$ with values in $\mathbb{R}$. Then, assuming that $r \geq 2$, $\hat{\boldsymbol{\theta}}$ satisfies (4) with $\Psi(\cdot|\boldsymbol{x}_\star)$ given by (8). Combining this with the fact that $F_{\boldsymbol{\theta}}^{-1}\left(1 - (1 - F_{\boldsymbol{\theta}}(R_r))^{\frac{s-1}{r}}\right) = R_r + \sigma\ln\left(\frac{s-1}{r}\right)$ yields that $\pi_{\text{MOLP}}^{(s)} = R_r + \hat{\sigma}_{\text{MOL}} \ln((s-1)/r)$ is the unique maximum observed likelihood predictor of R_s based on $\boldsymbol{R}_\star$, where the PMOLE of σ takes the form $\hat{\sigma}_{\text{MOL}} = \frac{1}{r}\sum_{i=1}^{r-1}(R_r - R_i)$. □

Remark 3.3 The PMOLE and the MLE of σ coincide. For the MLE of σ, we refer to Arnold et al. (1998, p. 127) (see also Remark 3.4).

Since extreme-value record values are generated from a location-scale family, it does not come as a surprise that linear prediction of extreme-value record values has already been treated (cf. Arnold et al. 1998, Example 5.6.3, p. 152, and section 5.6.2). However, it appears that maximum likelihood prediction of extreme-value record values has not been considered in the literature so far. The following result contains the expression of the MLP of R_s based on $\boldsymbol{R}_\star$.

Proposition 3.5 *Let $R_1, \ldots, R_s$ be the first s, $s \geq 3$, record values in a sequence of i.i.d. extreme-value random variables. For $r \in \mathbb{N}$, $2 \leq r < s$, the unique MLP of R_s and the unique PMLEs of μ and σ based on $\boldsymbol{R}_\star$ are given by*

$$\pi_{\text{MLP}}^{(s)} = R_r + \hat{\sigma}_{\text{ML}} \ln\left(\frac{s-1}{r}\right),$$

$$\hat{\sigma}_{\text{ML}} = \frac{1}{r+1} \sum_{i=1}^{r-1}(R_r - R_i) \quad \text{and} \quad \hat{\mu}_{\text{ML}} = R_r + \hat{\sigma}_{\text{ML}} \ln\left(\frac{s-1}{sr}\right),$$

respectively. If $s = r + 1$, the MLP takes the form $\pi_{\text{MOLP}}^{(s)} = R_r$.

Proof The predictive likelihood function of R_s and (μ, σ) given $\boldsymbol{R}_\star = \boldsymbol{x}_\star$, $\boldsymbol{x}_\star = (x_1, \ldots, x_r) \in \mathbb{R}_<$, satisfies

$$L_{\text{rv}}(x_s, \mu, \sigma \,|\, \boldsymbol{x}_\star) \propto \frac{1}{\sigma^{r+1}} \exp\left\{\frac{1}{\sigma}\sum_{i=1}^{r}(x_i - \mu) + \frac{x_s - \mu}{\sigma} - \exp\left\{\frac{x_s - \mu}{\sigma}\right\}\right\}$$

$$\times \left(\exp\left\{\frac{x_s - \mu}{\sigma}\right\} - \exp\left\{\frac{x_r - \mu}{\sigma}\right\}\right)^{s-r-1},$$

$x_s \geq x_r$, $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$. Observe that $PL(x_s, \mu, \sigma \,|\, \boldsymbol{x}_\star) \propto G(x_s, \mu, \sigma)H(x_s, \sigma)$, where $G(x_s, \mu, \sigma) = \exp\{-s\mu/\sigma - \exp\{(x_s - \mu)/\sigma\}\}$ and $H(x_s, \sigma) = \frac{1}{\sigma^{r+1}} \cdot \exp\left\{\frac{1}{\sigma}\sum_{i=1}^{r} x_i + x_s/\sigma\right\} (\exp\{x_s/\sigma\} - \exp\{x_r/\sigma\})^{s-r-1}$. Note that for any fixed $x_s \in [x_r, \infty)$ and $\sigma \in \mathbb{R}_+$, we have that $\lim_{\mu \to \pm\infty} G(x_s, \mu, \sigma) = 0$. As $\mu \mapsto G(x_s, \mu, \sigma)$ is a continuous function, it possesses a global maximum in $\mathbb{R}$. Moreover, we have that $\partial G(x_s, \mu, \sigma)/\partial\mu = 0$ if and only if $\mu = x_s - \sigma \ln(s)$.

Hence, the function $\mu \mapsto G(x_s, \mu, \sigma)$, $\mu \in \mathbb{R}$, attains its global maximum value at a unique point and the global maximum value equals

$$G(x_s, x_s - \sigma \ln(s), \sigma) = \exp\left\{-\frac{s}{\sigma}x_s + s \ln(s) - s\right\}.$$

Observe that the function $G(x_s, x_s - \sigma \ln(s), \sigma) \propto e^{-\frac{s}{\sigma}x_s}$. Consequently, it suffices to show that

$$J(x_s, \sigma) = H(x_s, \sigma) \exp\left\{-\frac{s}{\sigma}x_s\right\} = \frac{1}{\sigma^{r+1}} \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{r-1}(x_r - x_i)\right\}$$

$$\times \left(\exp\left\{-\frac{x_s - x_r}{\sigma}\right\}\right)^{r} \left(1 - \exp\left\{-\frac{x_s - x_r}{\sigma}\right\}\right)^{s-r-1}$$

attains its maximum uniquely in $[x_r, \infty) \times \mathbb{R}_+$. For fixed $\sigma \in \mathbb{R}_+$, the function $x_s \mapsto \left(e^{-\frac{x_s - x_r}{\sigma}}\right)^{r} \left(1 - e^{-\frac{x_s - x_r}{\sigma}}\right)^{s-r-1}$, $x_s \in [x_r, \infty)$, attains its global maximum value at the unique point $x_s = x_r + \sigma \ln(\frac{s-1}{r})$. Moreover, we have

that $J(x_r + \sigma \ln(\frac{s-1}{r}), \sigma) \propto \frac{1}{\sigma^{r+1}} e^{-\frac{1}{\sigma} \sum_{i=1}^{r-1}(x_r - x_i)}$. Finally, the function $\sigma \mapsto \frac{1}{\sigma^{r+1}} e^{-\frac{1}{\sigma} \sum_{i=1}^{r-1}(x_r - x_i)}$, $\sigma \in \mathbb{R}_+$, attains its global maximum value at a unique point, which is given by $\sigma = \frac{1}{r+1} \sum_{i=1}^{r-1}(x_r - x_i)$. Combing all these results completes the proof. $\qquad\square$

Remark 3.4 Observe that neither the PMLE of μ nor the PMLE of σ coincides with the MLE of the corresponding parameter. The MLEs of μ and σ are given by $\hat{\mu} = R_r + \hat{\sigma} \ln(r)$ and $\hat{\sigma} = \frac{1}{r} \sum_{i=1}^{r-1}(R_r - R_i)$. Their derivation can be found, e.g., in Arnold et al. (1998, p. 127). Moreover, the PMLE of μ depends on which future record value is to be predicted, as the appearance of the index of the future record value in the expression for $\hat{\mu}_{ML}$ reveals.

3.2.1 Comparison based on the MSE

In preparation for the comparison of the MOLP to the MLP, we derive their MSEs. In what follows, for $r, s \in \mathbb{N}, r < s$, the following notation will be used:

$$\alpha_{r,s}(1) = \sum_{i=r}^{s-1} \frac{1}{i}, \quad \alpha_{r,s}(2) = \sum_{i=r}^{s-1} \frac{1}{i^2}.$$

Lemma 3.4 *For $s \geq 3$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. extreme-value random variables. The MSEs of the MOLP and the MLP of R_s based on $\mathbf{R}_\star$ are given by*

$$\frac{MSE(\pi_{MOLP}^{(s)})}{\sigma^2} = \alpha_{r,s}(2) + \frac{\alpha_{r,s}(1)^2}{r} + \frac{r-1}{r}\left(\alpha_{r,s}(1) - \ln\left(\frac{s-1}{r}\right)\right)^2$$

and

$$\frac{MSE(\pi_{MLP}^{(s)})}{\sigma^2} = \alpha_{r,s}(2) + \left(\alpha_{r,s}(1) - \ln\left(\frac{s-1}{r+1}\right)\frac{r-1}{r+1}\right)^2 + \ln^2\left(\frac{s-1}{r}\right)\frac{r-1}{r+1},$$

respectively.

Proof Since $R_i - R_{i-1} \sim \mathcal{E}xp(\sigma/(i-1)), i \geq 2$,

$$E(\hat{\sigma}_{MOL}) = E\left(\frac{1}{r}\sum_{i=1}^{r-1}(R_r - R_i)\right) = E\left(\frac{1}{r}\sum_{i=1}^{r-1}\sum_{j=i}^{r-1}(R_{j+1} - R_j)\right)$$

$$= E\left(\frac{1}{r}\sum_{j=1}^{r-1} j(R_{j+1} - R_j)\right) = \sigma\frac{r-1}{r}.$$

Combining this with the expression for the expected value of an extreme-value record value (see Arnold et al. 1998, p. 32), the claimed expression for the bias of the MOLP

readily follows. As for the MSE of the MOLP, observe first that by the fact that $R_i - R_{i-1} \overset{\text{i.i.d.}}{\sim} \mathcal{E}xp(\sigma/(i-1))$, $i \geq 2$, we have $r\hat{\sigma}_{\text{MOL}} \sim \mathcal{G}amma(r-1, \sigma)$, where $\mathcal{G}amma(a, b)$ denotes the gamma distribution with shape parameter $a \in \mathbb{R}_+$ and scale parameter $b \in \mathbb{R}_+$. Consequently, $E((\hat{\sigma}_{\text{MOL}})^2) = \sigma^2(r-1)/r$. Hence, using again the independence of spacings, we obtain

$$\frac{\text{MSE}(\pi_{\text{MOLP}}^{(s)})}{\sigma^2} = \frac{1}{\sigma^2} E((R_s - R_r)^2) - \frac{2}{\sigma^2} \ln\left(\frac{s-1}{r}\right) E((R_s - R_r)\hat{\sigma}_{\text{MOL}})$$

$$+ \frac{1}{\sigma^2} \ln^2\left(\frac{s-1}{r}\right) E((\hat{\sigma}_{\text{MOL}})^2)$$

$$= \alpha_{r,s}(2) + \frac{\alpha_{r,s}(1)^2}{r} + \frac{r-1}{r}\left(\alpha_{r,s}(1) - \ln\left(\frac{s-1}{r}\right)\right)^2.$$

Given that $\hat{\sigma}_{\text{ML}} = (r/(r+1))\hat{\sigma}_{\text{MOL}}$, the derivation of $\text{MSE}(\pi_{\text{MLP}}^{(s)})$ proceeds along the same lines. $\qquad\square$

Proposition 3.6 *For $s \geq 3$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. extreme-value random variables. For $r \in \mathbb{N}$, $2 \leq r \leq s-2$, let $\pi_{\text{MOLP}}^{(s)}$ and $\pi_{\text{MLP}}^{(s)}$ be, respectively, the MOLP and the MLP of R_s based on $\mathbf{R}_\star$. Then,*

$$\text{MSE}(\pi_{\text{MOLP}}^{(s)}) < \text{MSE}(\pi_{\text{MLP}}^{(s)}).$$

Proof We have

$$\frac{\text{MSE}(\pi_{\text{MLP}}^{(s)}) - \text{MSE}(\pi_{\text{MOLP}}^{(s)})}{\sigma^2}$$

$$= \ln\left(\frac{s-1}{r}\right)\frac{r-1}{r(r+1)}\left\{2\left(\alpha_{r,s}(1) - \ln\left(\frac{s-1}{r}\right)\right) + \frac{1}{r+1}\ln\left(\frac{s-1}{r}\right)\right\}. \tag{9}$$

The function $x \mapsto x^{-1}$ is decreasing on $\mathbb{R}_+$, and $\ln((s-1)/r) = \int_r^{s-1} x^{-1}dx$. Thus, we have that $\alpha_{r,s}(1) - \ln((s-1)/r) > 0$. Hence, since the factor in front of the curly brackets and the second summand between the curly brackets are obviously positive, the assertion follows. $\qquad\square$

Figure 3a contains the contour plot of the relative efficiency of $\pi_{\text{MLP}}^{(s)}$ relative to $\pi_{\text{MOLP}}^{(s)}$ of R_s based on $\mathbf{R}_\star$ for all valid combinations of r and s in the range of 2–200. Table 3 contains relative efficiencies of the predictors for selected r and s. The numerical results confirm the superiority of the MOLP over the MLP in terms of mean squared error. Though, the gains in efficiency are smaller than in the case of exponential record values. The highest efficiencies are achieved for small values of r, and the efficiency gains quickly become negligible as r increases. Still, from the perspective of practical applications, where rather small numbers of observed record values are common, this fact makes the MOLP an attractive alternative to the MLP.
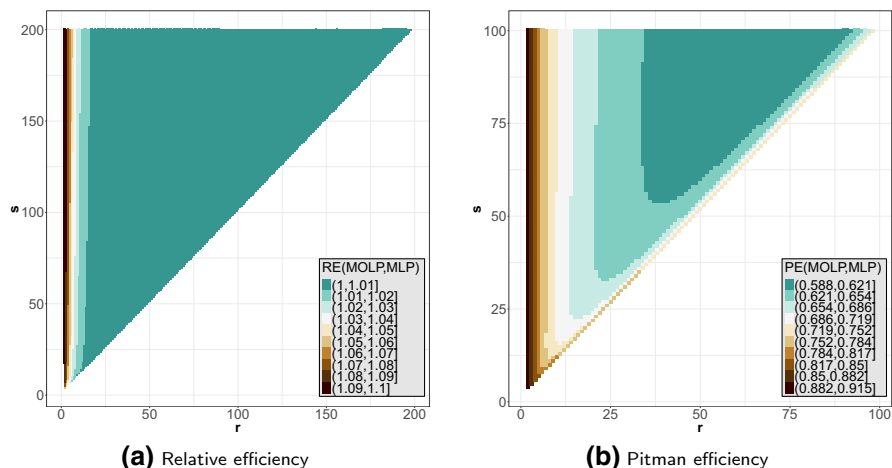
(a) Relative efficiency

(b) Pitman efficiency

Fig. 3 Contour plots of the relative efficiency RE(MOLP, MLP) = $\text{MSE}(\pi_{\text{MLP}}^{(s)})/\text{MSE}(\pi_{\text{MOLP}}^{(s)})$ and the Pitman efficiency PE(MOLP, MLP) = $P(|R_s - \pi_{\text{MOLP}}^{(s)}| < |R_s - \pi_{\text{MLP}}^{(s)}|)$ of the MLP of R_s relative to the MOLP of R_s based on extreme-value record values $\boldsymbol{R}_\star$ for all valid combinations of r and s in the range 2–200 with $s = r + 1$ omitted

Table 3 Relative efficiencies RE(MOLP, MLP) = $\text{MSE}(\pi_{\text{MLP}}^{(s)})/\text{MSE}(\pi_{\text{MOLP}}^{(s)})$ of the MLP of R_s relative to the MOLP of R_s based on extreme-value record values $\boldsymbol{R}_\star$ for selected r and s

r	s						
	$r+2$	$r+3$	$r+4$	$r+5$	$r+10$	$r+15$	$r+20$
2	1.053	1.067	1.073	1.076	1.083	1.085	1.087
5	1.023	1.031	1.035	1.037	1.042	1.045	1.047
10	1.008	1.010	1.012	1.012	1.014	1.015	1.016
15	1.004	1.005	1.006	1.006	1.007	1.007	1.008
20	1.002	1.003	1.003	1.004	1.004	1.004	1.004
25	1.001	1.002	1.002	1.002	1.003	1.003	1.003
50	1.000	1.001	1.001	1.001	1.001	1.001	1.001
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.2.2 Comparison based on Pitman's measure of closeness

We compare the predictors based on Pitman's measure of closeness. Similarly to Lemma 3.3, we use the same arguments as in Nagaraja (1986, p. 14) to compute the Pitman efficiency. The corresponding result is contained in the following Lemma. Below, $\mathcal{HE}xp(\alpha_1, \ldots, \alpha_n)$ denotes the hypoexponential distribution with pairwise different rate parameters $\alpha_1, \ldots, \alpha_n \in \mathbb{R}_+$ (see Ross 2014).

Lemma 3.5 *Let* $R_1, \ldots, R_s$ *be the first* s, $s \geq 3$, *record values in a sequence of i.i.d. extreme-value random variables. Let* $\pi_{\text{MOLP}}^{(s)}$ *and* $\pi_{\text{MLP}}^{(s)}$ *be, respectively, the MOLP*

and the MLP of R_s based on $\boldsymbol{R}_\star$, $2 \leq r \leq s - 2$. Then,

$$P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|) = P\left(\frac{U}{T} > \frac{2r + 1}{2r(r + 1)} \ln\left(\frac{s - 1}{r}\right)\right).$$

Here, U and T are independent, $U \sim \mathcal{HExp}(r, \ldots, (s - 1))$, $T \sim \mathcal{Gamma}(r - 1, 1)$.

Remark 3.5 Observe that U/T from Proposition 3.5 is distributed as the ratio of two hypoexponential random variables. Hence, using Kadri and Smaili (2014, Theorem 1), we obtain a representation of the cdf of U/T. For $x \in \mathbb{R}_+$,

$$F^{U/T}(x) = (s - 1)\binom{s - 2}{r - 1}\sum_{i=r}^{s-1}(-1)^{i-r}\binom{s - r - 1}{i - r}i^{-1}I_{\frac{x}{i^{-1}+x}}(1, r - 1), \quad (10)$$

where $I_.(a, b)$ is the regularized incomplete beta function with parameters a, $b > 0$. A more compact representation of $F^{U/T}$ can be achieved if one observes that the sum in (10) is an alternating binomial sum. More specifically, we have that for $x \in \mathbb{R}_+$,

$$\sum_{i=r}^{s-1}(-1)^{i-r}\binom{s - r - 1}{i - r}i^{-1}I_{\frac{x}{i^{-1}+x}}(1, r - 1)$$

$$= (-1)^{s-r-1}\sum_{i=0}^{s-r-1}(-1)^{s-r-1-i}\binom{s - r - 1}{i}f_{r,x}(i),$$

where for $i \in \mathbb{N} \cup \{0\}$, $f_{r,x}(i) = \frac{1}{i+r}I_{\frac{x}{(i+r)^{-1}+x}}(1, r - 1)$. Using the fact that iterated forward differences can be expressed by alternating binomial sums, we obtain a compact representation of $F^{U/T}$ as

$$F^{U/T}(x) = (-1)^{s-r-1}(s - 1)\binom{s - 2}{r - 1}\Delta^{s-r-1}f_{r,x}, \quad x \in \mathbb{R}_+,$$

where the $(s - r - 1)$th fold difference is to be computed for $i = 0$. Thus, the Pitman efficiency $P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|)$ of the MLP relative to the MOLP can be expressed as

$$P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|)$$

$$= 1 - (-1)^{s-r-1}(s - 1)\binom{s - 2}{r - 1}\Delta^{s-r-1}f_{r,x}|_{x = \frac{2r+1}{2r(r+1)}\ln(\frac{s-1}{r})},$$

with the $(s - r - 1)$th forward difference computed for $i = 0$. Since alternating sums can be numerically problematic, for an efficient and accurate procedure to compute $P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|)$, it is advisable to use high-precision arithmetic. See `sumBinomMpfr()` in R package `Rmpfr` and its documentation.

Table 4 Pitman efficiencies PE(MOLP, MLP) $= P(|R_s - \pi_{\mathrm{MOLP}}^{(s)}| < |R_s - \pi_{\mathrm{MLP}}^{(s)}|)$ of the MLP of R_s relative to the MOLP of R_s based on extreme-value record values $\boldsymbol{R}_\star$ for selected r and s

r	s						
	$r+2$	$r+3$	$r+4$	$r+5$	$r+10$	$r+15$	$r+20$
2	0.915	0.909	0.908	0.909	0.910	0.912	0.914
5	0.827	0.802	0.794	0.791	0.791	0.792	0.794
10	0.786	0.748	0.732	0.725	0.715	0.715	0.716
15	0.770	0.726	0.707	0.697	0.681	0.679	0.678
20	0.762	0.715	0.693	0.682	0.661	0.657	0.656
25	0.757	0.707	0.685	0.672	0.649	0.643	0.641
50	0.747	0.692	0.667	0.651	0.620	0.610	0.606
80	0.743	0.687	0.660	0.643	0.609	0.596	0.590

Figure 3b contains the contour plot of the Pitman efficiency of $\pi_{\mathrm{MLP}}^{(s)}$ relative to $\pi_{\mathrm{MOLP}}^{(s)}$ for all valid combinations of r and s in the range 2–100. Table 4 contains Pitman efficiencies of the predictors, which were computed using the expression derived in Remark 3.5, for selected r and s. Remarkably, the Pitman efficiencies are almost identical to those of the MLP relative to the MOLP for exponential record values presented in Table 2 (unknown μ).

3.3 Power-function distribution

In the previous two subsections, we have demonstrated by the examples of the exponential and extreme-value distributions the simplicity of deriving the MOLP as well as its superior performance over the MLP. With the power-function distribution, the situation is different. The MOLP can be shown to uniquely exist, though its computation is rather laborious. Moreover, it turns out that the MLP does not exist.

In what follows, we assume that $(R_n)_{n \in \mathbb{N}}$ is the sequence of record values in a sequence of i.i.d. power-function random variables. The density, cumulative distribution and quantile functions of the power-function distribution $\mathcal{P}ow(\theta)$ (or $\mathcal{B}eta(\theta, 1)$) with shape parameter $\theta \in \mathbb{R}_+$ are given by $f_\theta(x) = \theta x^{\theta-1}$, $F_\theta(x) = x^\theta$, $x \in (0, 1)$ and $F_\theta^{-1}(x) = x^{1/\theta}$, $x \in [0, 1)$. Next, for $r, s \in \mathbb{N}$, $r < s - 1$, we derive the MOLP of R_s based on $\boldsymbol{R}_\star = (R_1, \ldots, R_r)$. We also show that the MLP does not exist in the present situation.

Proposition 3.7 *For $s \geq 3$, let $R_1, \ldots, R_s$ be the first s record values in a sequence of i.i.d. power-function random variables. For $r \in \mathbb{N}$, $1 \leq r < s - 1$, the unique MOLP of R_s based on $\boldsymbol{R}_\star$ is given by*

$$\pi_{\mathrm{MOLP}}^{(s)} = \left(1 - \left(1 - R_r^{\hat{\theta}_{\mathrm{MOL}}}\right)^{\frac{s-1}{r}}\right)^{1/\hat{\theta}_{\mathrm{MOL}}},$$

where $\hat{\theta}_{\mathrm{MOL}}$ is the unique PMOLE of θ and is obtained as the unique positive solution of the equation

$$\ln\left(\prod_{i=1}^{r} R_i\right) + \frac{r}{\theta} + \sum_{i=1}^{r} \frac{R_i^\theta \ln(R_i)}{1 - R_i^\theta} + \frac{r}{\ln(1 - R_r^\theta)} \frac{R_r^\theta \ln(R_r)}{1 - R_r^\theta} = 0$$

with respect to θ.

Proof We present a sketch of the proof. The details can be found in Volovskiy (2018, Section 5.3.6). With the above choice for f_θ and F_θ, the function $\Psi(\cdot|\boldsymbol{x}_\star), \boldsymbol{x}_\star \in (0, 1)^r_<,$ in (3) becomes

$$\Psi(\theta|\boldsymbol{x}_\star) = \theta^r \left(\prod_{i=1}^{r} x_i\right)^{\theta - 1} \left(-\frac{1}{\ln(1 - x_r^\theta)}\right)^r / \prod_{i=1}^{r}(1 - x_i^\theta), \quad \theta \in \mathbb{R}_+.$$

Fix some $\boldsymbol{x}_\star = (x_1, \ldots, x_r) \in (0, 1)^r_<$ and let f be the function $\Psi(\cdot|\boldsymbol{x}_\star)$. Using L'Hospital's rule, one shows that $\lim_{\theta \to 0} f(\theta) = \lim_{\theta \to \infty} f(\theta) = 0$. Since f is continuous, the preceding results imply that f possesses a global maximum. Next, we set $g(\theta) = \ln(f(\theta)), \theta \in \mathbb{R}_+$. The second derivative of g takes the form

$$g''(\theta) = -\frac{r}{\theta^2} + \sum_{i=1}^{r} \frac{x_i^\theta \ln^2(x_i)}{(1 - x_i^\theta)^2} + \frac{r}{\ln^2(1 - x_r^\theta)} \frac{x_r^{2\theta} \ln^2(x_r)}{(1 - x_r^\theta)^2} + \frac{r}{\ln(1 - x_r^\theta)} \frac{x_r^\theta \ln^2(x_r)}{(1 - x_r^\theta)^2}. \tag{11}$$

We show that g' is a strictly decreasing function, i.e., $g''(\theta) < 0, \theta \in \mathbb{R}_+$. Let $h_1(\theta)$ and $h_2(\theta)$ be equal, respectively, to the sum of the first two and the last two terms in (11). Using the inequality $x < -\ln(1 - x), x \in (0, 1)$, we infer that $h_2(\theta) < 0, \theta \in \mathbb{R}_+$. Next, note that $h_1(\theta) = (\sum_{i=1}^{r} f_{x_i}(\theta) - 1)/\theta^2$, where we have set $f_{x_i}(\theta) = x_i^\theta \ln^2(x_i^\theta)/(1 - x_i^\theta)^2, \theta \in \mathbb{R}_+$. Applying L'Hospital's rule twice, we infer that $\lim_{\theta \to 0} f_{x_i}(\theta) = 1, i = 1, \ldots, r$. We shall prove that each $f_{x_i}, i = 1, \ldots, r$, is strictly decreasing. Indeed, taking the logarithmic derivative of f_{x_i}, we obtain $\frac{d}{d\theta} \ln(f_{x_i}(\theta)) = \frac{1}{\theta}\left(2 + (1 + x_i^\theta) \ln(x_i^\theta)/(1 - x_i^\theta)\right)$. Thus, it suffices to show that $-2\frac{1-x}{1+x} > \ln(x)$, $x \in (0, 1)$, or, equivalently, $-\frac{2x}{2-x} > \ln(1 - x), x \in (0, 1)$. The Taylor series for $x \mapsto \ln(1 - x), x \in (-1, 1)$, and $x \mapsto -\frac{2x}{2-x}, x \in (-1, 1)$, at $x = 0$ are given by

$$\ln(1 - x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}, \ x \in (-1, 1) \quad \text{and} \quad -\frac{2x}{2 - x} = -\sum_{n=1}^{\infty} \frac{x^n}{2^{n-1}}, \ x \in (-1, 1).$$

Obviously, the difference of the first and the second power series has nonnegative coefficients. This implies the desired inequality. Consequently, $h_1(\theta) < 0, \theta \in \mathbb{R}_+$. This completes the proof of the assertion. □

Proposition 3.8 *The predictive likelihood function does not possess a global maximum. Hence, the MLP of R_s based on $\boldsymbol{R}_\star$ does not exist.*
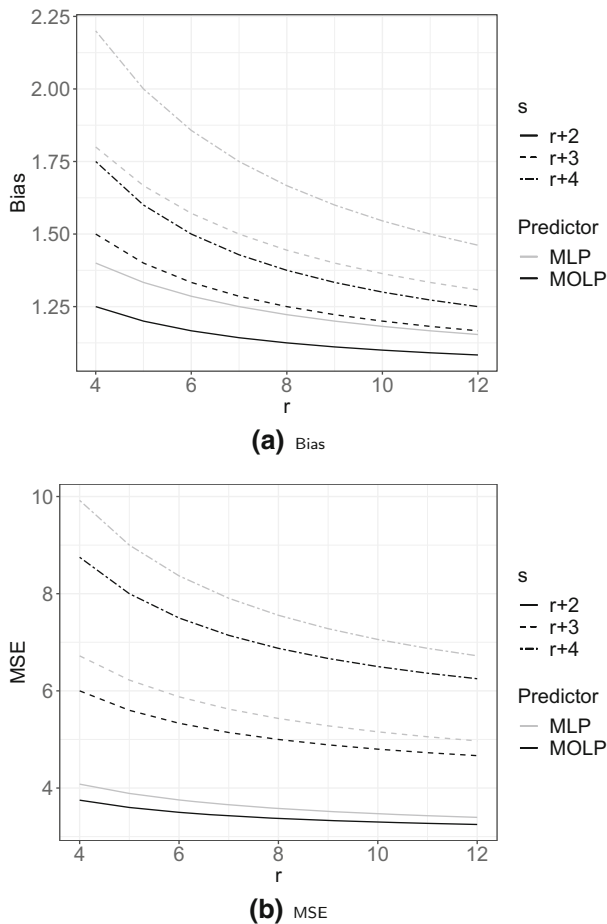
(a) Bias

(b) MSE

Fig. 4 Line plots of the bias and the MSE of MLP and MOLP of R_s based on two-parameter exponential record values $\boldsymbol{R}_\star$ for $r \in \{4, \ldots, 12\}$ and $s = r + 2, r + 3, r + 4$

Proof For $\boldsymbol{x}_\star \in (0, 1)_<^r$, the PLF satisfies

$$L_{\mathrm{rv}}(x_s, \theta | \boldsymbol{x}_\star) \propto \theta^{r+1} \left(\prod_{i=1}^{r} \frac{x_i^{\theta-1}}{1 - x_i^\theta} \right) \left(-\ln \left(\frac{1 - x_s^\theta}{1 - x_r^\theta} \right) \right)^{s-r-1} x_s^{\theta-1},$$

$x_s \in (x_r, 1)$, $\theta \in \mathbb{R}_+$. For any fixed $\theta \in \mathbb{R}_+$, $\lim_{x_s \to 1} L_{\mathrm{rv}}(x_s, \theta | \boldsymbol{x}_\star) = \infty$. Hence, no global maximum exists. □

4 Illustration

The prediction of future record values is illustrated for exponential distributions as in Sect. 3.1. In the literature, there are several real datasets of record values; an underlying exponential distribution is assumed by Dunsmore (1983) for data from a rock crushing machine (see also Awad and Raqab 2000) and by Razmkhah and Ahmadi (2013) for annual flood loss data. In the particular situation of an exponential distribution with unknown location parameter μ and scale parameter $\sigma > 0$, bias and mean squared prediction error of both, MLP and MOLP, can be stated explicitly. By noting that $E(R_r) = \mu + \sigma r, r \in \mathbb{N}$ (see Arnold et al. 1998), we find for $s > r + 1$:

$$E(R_s - \pi_{\text{MOLP}}^{(s)}) = \sigma \frac{s-1}{r},$$

$$E(R_s - \pi_{\text{MLP}}^{(s)}) = \sigma \left(\frac{2s}{r+1} - 1 \right),$$

such that both predictors are downward-biased. It is directly seen that the MOLP always has a smaller bias than the MLP. The mean squared prediction errors $\text{MSE}(\pi_{\text{MOLP}}^{(s)})$ and $\text{MSE}(\pi_{\text{MLP}}^{(s)})$ are stated in Sect. 3.1.1. Figure 4a, b shows the bias (in units of σ) and the MSE (in units of σ^2) of both predictors for several values of the number r of observations and $s = r + 2, r + 3, r + 4$.

5 Conclusion

Based on the observed predictive likelihood function first studied as a tool for deriving predictive inferences by Bayarri et al. (1987), a novel likelihood-based prediction procedure is proposed. The prediction method is successfully applied to the problem of future (upper) record value prediction, and for underlying exponential and extreme-value distributions, it is demonstrated that the resulting predictors exhibit superior performance relative to predictors produced by the widely applied maximum likelihood prediction procedure. The obtained predictors will be useful in reliability applications involving modeling repairable systems and, more generally, in areas where the underlying stochastic dynamics are adequately described by nonhomogeneous Poisson processes.

References

Ahmadi J, Doostparast M (2006) Bayesian estimation and prediction for some life distributions based on record values. Stat Pap 47(3):373–392

Ahsanullah M (1980) Linear prediction of record values for the two parameter exponential distribution. Ann Inst Stat Math 32(1):363–368

Arnold BC, Balakrishnan N, Nagaraja HN (1998) Records. Wiley, Hoboken

Awad AM, Raqab MZ (2000) Prediction intervals for the future record values from exponential distribution: comparative study. J Stat Comput Simul 65(1–4):325–340

Balakrishnan N, Cramer E (2014) The art of progressive censoring. Springer, New York

Basak P, Balakrishnan N (2003) Maximum likelihood prediction of future record statistic. In: Lindqvist B, Doksum K (eds) Mathematical and statistical methods in reliability, chap 11. World Scientific Publishing, Singapore, pp 159–175

Bayarri MJ, DeGroot MH (1988) Discussion: auxiliary parameters and simple likelihood functions. In: Berger JO, Wolpert RL (eds) The likelihood principle, 2nd edn. Institute of Mathematical Statistics, Hayward, pp 160.3–160.7

Bayarri MJ, DeGroot MH, Kadane JB (1987) What is the likelihood function? (with discussion). In: Gupta SS, Berger JO (eds) Statistical decision theory and related topics IV. Springer, New York, pp 3–16

Berger JO, Wolpert RL (1988) The likelihood principle, 2nd edn. Institute of Mathematical Statistics, Hayward

Birnbaum A (1962) On the foundations of statistical inference. J Am Stat Assoc 57(298):269–306

Bjørnstad JF (1996) On the generalization of the likelihood function and the likelihood principle. J Am Stat Assoc 91(434):791–806

Burkschat M (2009) Linear estimators and predictors based on generalized order statistics from generalized Pareto distributions. Commun Stat Theory Methods 39(2):311–326

Chandler KN (1952) The distribution and frequency of record values. J R Stat Soc Ser B Stat Methodol 14:220–228

Dunsmore IR (1983) The future occurrence of records. Ann Inst Stat Math 35(1):267–277

Dziubdziela W, Kopociński B (1976) Limiting properties of the k-th record values. Appl Math (Warsaw) 2(15):187–190

Groeneveld RA, Meeden G (1977) The mode, median, and mean inequality. Am Stat 31(3):120–121

Gupta RC, Kirmani SNUA (1988) Closure and monotonicity properties of nonhomogeneous Poisson processes and record values. Probab Eng Inf Sci 2(4):475–484

Kadri T, Smaili K (2014) The exact distribution of the ratio of two independent hypoexponential random variables. Br J Math Comput Sci 4(18):2665–2675

Kaminsky KS (1987) Prediction of ibnr claim counts by modelling the distribution of report lags. Insur Math Econ 6(2):151–159

Kaminsky KS, Rhodin LS (1985) Maximum likelihood prediction. Ann Inst Stat Math 37(3):507–517

Kamps U (1995) A concept of generalized order statistics. J Stat Plan Inference 48(1):1–23

Keseling C (1999) Charakterisierung von Wahrlichkeitsverteilungen durch verallgemeinerte Ordnungsstatistiken. Ph.D. thesis, RWTH Aachen University

Madi MT, Raqab MZ (2004) Bayesian prediction of temperature records using the Pareto model. Environmetrics 15(7):701–710

Nadar M, Kızılaslan F (2015) Estimation and prediction of the Burr type XII distribution based on record values and inter-record times. J Stat Comput Simul 85(16):3297–3321

Nagaraja HN (1986) Comparison of estimators and predictors from two-parameter exponential distribution. Sankhyā Ser B 48(1):10–18

Nayak TK, Kundu S (2002) Some remarks on generalizations of the likelihood function and the likelihood principle. In: Balakrishnan N (ed) Advances on theoretical and methodological aspects of probability and statistics. Taylor and Francis, London, pp 1999–2012

Nevzorov VB (2001) Records: mathematical theory. American Mathematical Society, Providence

Raqab MZ (2001) Optimal prediction intervals for the exponential distribution based on generalized order statistics. IEEE Trans Reliab 50(1):112–115

Raqab MZ (2007) Exponential distribution records: different methods of prediction. In: Ahsanullah M, Raqab MZ (eds) Recent developments in ordered random variables, chap 16. Nova Science Publishers, Hauppauge, pp 239–251

Raqab MZ, Ahmadi J, Doostparast M (2007) Statistical inference based on record data from Pareto model. Statistics 41(2):105–118

Razmkhah M, Ahmadi J (2013) Pitman closeness of current k-records to population quantiles. Stat Probab Lett 83(1):148–156

Ross SM (2014) Introduction to probability models. Academic Press, Oxford

Volovskiy G (2018) Likelihood-based prediction in models of ordered data. Ph.D. thesis, RWTH Aachen University