

Article

Objective Bayesian Prediction of Future Record Statistics Based on the Exponentiated Gumbel Distribution: Comparison with Time-Series Prediction

Yongku Kim ¹  and Jung In Seo ^{2,*}¹ Department of Statistics, Kyungpook National University, Daegu 41566, Korea; kim.1252@knu.ac.kr² Division of Convergence Education, Halla University, Wonju 26404, Korea

* Correspondence: ji.seo@halla.ac.kr or leehoo1928@gmail.com

Received: 13 July 2020; Accepted: 2 August 2020; Published: 1 September 2020



Abstract: The interest in the study of record statistics has been increasing in recent years in the context of predicting stock markets and addressing global warming and climate change problems such as cyclones and floods. However, because record values are mostly rare observed, its probability distribution may be skewed or asymmetric. In this case, the Bayesian approach with a reasonable choice of the prior distribution can be a good alternative. This paper presents an objective Bayesian method for predicting future record values when observed record values have a two-parameter exponentiated Gumbel distribution with the scale and shape parameters. For objective Bayesian analysis, objective priors such as the Jeffreys and reference priors are first derived from the Fisher information matrix for the scale and shape parameters, and an analysis of the resulting posterior distribution is then performed to examine its properness and validity. In addition, under the derived objective prior distributions, a simple algorithm using a pivotal quantity is proposed to predict future record values. To validate the proposed approach, it was applied to a real dataset. For a closer examination and demonstration of the superiority of the proposed predictive method, it was compared to time-series models such as the autoregressive integrated moving average and dynamic linear model in an analysis of real data that can be observed from an infinite time series comprising independent sample values.

Keywords: exponentiated gumbel distribution; objective Bayesian analysis; record value; time series

1. Introduction

The occurrence of extreme events such as extreme temperatures, excess flood peaks, and rapid increases in pollutant concentrations has steadily increased over the past decade. However, the volume of data generated by such events is relatively small compared to big data generated by normal events occurring daily. In this case, providing accurate predictions for extreme events might be a challenging task due to the insufficient number of past records. Bayesian methods may be more suitable for modeling data to small sample sizes with skewness or lack of symmetry as in the case of record values, provided that the Bayesian method involves a reasonable choice for the prior distribution because the Bayesian methods do not rely on asymptotic theory in the same way that frequentist methods do.

The concept of record values was introduced by Chandler [1] and it can be described as follows. Let $\{X_1, X_2, \dots\}$ be a sequence of independent and identically distributed random variables. Then, for every $i < j$, X_j is called the upper (lower) record value if $X_j > (<) X_i$, which indicates that X_j is higher (lower) than all previous observations. That is, the upper (lower) record values include the members of a series that are larger (smaller) than all preceding members. The indices for which upper

record values occur are informed by the record times $\{U(k), k \geq 1\}$, where $U(k) = \min\{j | j > U(k-1), X_j > X_{U(k-1)}\}$, $k > 1$, with $U(1) = 1$. The record times for the lower record values are $\{L(k), k \geq 1\}$, where $L(k) = \min\{j | j > L(k-1), X_j < X_{L(k-1)}\}$, $k > 1$, with $L(1) = 1$. Therefore, the sequences of the upper and lower record values are denoted as $\{X_{U(k)}, k = 1, 2, \dots\}$ and $\{X_{L(k)}, k = 1, 2, \dots\}$, respectively, from the original sequence $\{X_1, X_2, \dots\}$.

Since such record values arise in many real-world situations related to climate, economics, sports, and life test, relevant studies have been conducted in various fields. Coles and Tawn [2] analyzed a daily rainfall series for modeling the extremes of the rainfall process in the context of record values. Madi and Raqab [3] analyzed average temperatures in Neuenburg, Switzerland, using a Bayesian predictive method for record values from the Pareto distribution. Wergen et al. [4] analyzed both the probability of occurrence and PDF of record-breaking values for temperatures in Europe and the United States. Seo and Kim [5] proposed an objective Bayesian inference method for record values from the Gumbel distribution, which was applied to the concentration analysis of sulfur dioxide. Seo and Kang [6] proposed a estimation method using record values from an exponentiated half-logistic distribution in Bayesian and non-Bayesian perspectives. The authors demonstrated the efficiency of their method by comparing it to the existing estimation methods through rainfall data analysis.

This paper proposes a predictive method based on an objective Bayesian approach that can save the effort of finding an exact prior distribution when there is no sufficient information in the context of record statistics values from the exponentiated Gumbel distribution (EGD) with cumulative distribution function (CDF)

$$F(x) = \left(e^{-e^{-x/\sigma}}\right)^\lambda, \quad -\infty < x < \infty, \lambda, \sigma > 0, \quad (1)$$

where σ and λ denote the scale and shape parameters, respectively. The EGD is a generalized version of the GD, which is the most widely applied statistical distribution in the extreme value analysis of extreme events such as global warming, floods, heavy rainfall, and high wind speeds. The EGD can be considered as being simply the λ th power of the CDF of the GD with the scale parameter σ . Therefore, the EGD can lead to an improved performance and applicability of models built over a variety of complex datasets compared to the GD. Note that it is possible to apply time-series techniques if no data are lost during acquiring record values since the lower record time $L(k)$ is the serial number of record values in an infinite time series. For comparison with the proposed objective Bayesian predictive method, two types of time-series models are considered in this study: the autoregressive integrated moving average (ARIMA) model introduced by Box et al. [7] and the dynamic linear model (DLM) developed by West and Harrison [8].

The rest of the paper is organized as follows. Section 2 presents objective priors for unknown parameters of the EGD in the context of record statistics values, along with the corresponding posterior analysis and predictive method. Section 3 provides a brief description of the ARIMA and DLM that are employed in this study as benchmarks to validate the proposed objective Bayesian method. Section 4 presents the results of applying both the time-series and objective Bayesian models to real data. Section 5 concludes this paper.

2. Bayesian Prediction

The aim of this study is to predict future lower record values based on an objective Bayesian approach. To accomplish this, an objective Bayesian approach that does not require determining hyperparameters is presented first. The following subsection introduces objective priors based on the Fisher information (FI) matrix for unknown parameters of the EGD with the CDF (1).

2.1. Objective Prior

Let $X_{L(1)}, \dots, X_{L(k)}$ be the lower record values from the CDF (1). Then, the corresponding likelihood function and its natural logarithm can be expressed as

$$\begin{aligned} L(\lambda, \sigma) &= f(x_{L(k)}) \prod_{i=1}^{k-1} \frac{f(x_{L(i)})}{F(x_{L(i)})} \\ &= \left(\frac{\lambda}{\sigma}\right)^k e^{-\lambda e^{-x_{L(k)}/\sigma}} e^{-\sum_{i=1}^k x_{L(i)}/\sigma} \end{aligned}$$

and

$$\log L(\lambda, \sigma) = k \log \lambda - k \log \sigma - \lambda e^{-x_{L(k)}/\sigma} - \frac{1}{\sigma} \sum_{i=1}^k x_{L(i)}, \quad (2)$$

respectively. For computational convenience, let $\theta = 1/\sigma$. Then, based on the log-likelihood (2), the FI matrix for (λ, θ) can be defined as follows.

Proposition 1. The FI matrix for (λ, θ) is of the form

$$\mathbf{I}(\lambda, \theta) = \begin{pmatrix} k/\lambda^2 & kQ_1(\lambda)/(\lambda\theta) \\ kQ_1(\lambda)/(\lambda\theta) & kQ_2(\lambda)/\theta^2 \end{pmatrix}, \quad (3)$$

where

$$\begin{aligned} Q_1(\lambda) &= \log \lambda - \psi(k+1), \\ Q_2(\lambda) &= [Q_1(\lambda)]^2 + \psi_1(k+1) + 1, \end{aligned}$$

and $\psi(\cdot)$ and $\psi_1(\cdot)$ are the digamma and trigamma functions, respectively.

Proof. In the FI matrix

$$\mathbf{I}(\lambda, \theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix},$$

the element I_{11} can be easily computed, while the other elements can be expressed as

$$\begin{aligned} I_{12} &= E \left(-\frac{\partial^2}{\partial \lambda \partial \theta} \log L(\lambda, \theta) \right) \\ &= E \left(X_{L(k)} e^{-\theta X_{L(k)}} \right) \\ &= I_{21} \end{aligned}$$

and

$$\begin{aligned} I_{22} &= E \left(-\frac{\partial^2}{\partial \theta^2} \log L(\lambda, \theta) \right) \\ &= \frac{k}{\theta^2} + \lambda E \left(X_{L(k)}^2 e^{-\theta X_{L(k)}} \right). \end{aligned}$$

Then, the proof is completed given the marginal density function of $X_{L(i)}$ defined in Ahsanullah [9] as

$$f_{X_{L(i)}}(x) = \frac{1}{\Gamma(k)} [-\log F(x)]^{i-1} f(x)$$

and assuming $Y = \lambda e^{-\theta X_{L(k)}}$. \square

The objective priors such as the Jeffreys and reference priors based on the FI matrix (3) are defined according to the following theorem.

Theorem 1. *The Jeffreys prior for (λ, θ) is*

$$\pi_J(\lambda, \theta) \propto \frac{1}{\lambda\theta}.$$

Proof. According to the definition of the Jeffreys prior (Jeffreys [10]), it follows that

$$\begin{aligned}\pi_J(\lambda, \theta) &\propto \sqrt{|\mathbf{I}|} \\ &= \frac{k}{\lambda\theta} [\psi_1(k+1) + 1],\end{aligned}$$

where $|\mathbf{I}|$ denotes the determinant of the FI matrix (3). This completes the proof. \square

In the following, the reference priors for each parameter of interest are derived from the algorithm provided in Berger and Bernardo [11].

Theorem 2. *If λ is the parameter of interest, the reference prior for (λ, θ) is*

$$\pi_{R1}(\lambda, \theta) \propto \frac{1}{\theta\lambda\sqrt{Q_2(\lambda)}},$$

and, if θ is the parameter of interest, the reference prior for (θ, λ) is

$$\pi_{R2}(\theta, \lambda) \propto \frac{1}{\lambda\theta}.$$

Proof. When λ is the parameter of interest, the conditional prior distribution of θ given λ can be defined based on the FI matrix (3) as

$$\begin{aligned}\pi(\theta|\lambda) &= \sqrt{I_{22}} \\ &= \frac{1}{\theta} \sqrt{kQ_2(\lambda)}.\end{aligned}$$

Then, by choosing a sequence of compact sets $\Omega_i = (d_{1i}, d_{2i}) \times (d_{3i}, d_{4i})$ for (λ, θ) such that $d_{1i}, d_{3i} \rightarrow 0, d_{2i}, d_{4i} \rightarrow \infty$ as $i \rightarrow \infty$, it follows that

$$\begin{aligned}k_{1i}^{-1}(\lambda) &= \int_{d_{4i}}^{d_{3i}} \pi(\theta|\lambda) d\theta \\ &= \sqrt{kQ_2(\lambda)} [\log(d_{4i}) - \log(d_{3i})],\end{aligned}$$

and

$$\begin{aligned}p_i(\theta|\lambda) &= k_{1i}(\lambda) \pi(\theta|\lambda) \\ &= \frac{1}{\theta [\log(d_{4i}) - \log(d_{3i})]}.\end{aligned}\tag{4}$$

In addition, the marginal reference prior for λ can be defined based on the FI matrix (3) and (4) as

$$\begin{aligned}\pi_i(\lambda) &= \exp \left[\frac{1}{2} \int_{d_{3i}}^{d_{4i}} p_i(\theta|\lambda) \log \left(\frac{|\mathbf{I}|}{I_{22}} \right) d\theta \right] \\ &= \frac{1}{\lambda \sqrt{Q_2(\lambda)}},\end{aligned}$$

which leads to the following reference prior:

$$\begin{aligned}\pi_{R1}(\lambda, \theta) &= \lim_{i \rightarrow \infty} \left[\frac{k_{1i}(\lambda) \pi_i(\lambda)}{k_{1i}(\lambda_0) \pi_i(\lambda_0)} \right] \pi(\theta|\lambda) \\ &\propto \frac{1}{\theta \lambda \sqrt{Q_2(\lambda)}}\end{aligned}$$

for any fixed point λ_0 . When θ is the parameter of interest, the same argument is applied.

Let

$$\begin{aligned}\pi(\lambda|\theta) &= \sqrt{I_{11}} \\ &= \frac{\sqrt{k}}{\lambda}.\end{aligned}$$

Then,

$$\begin{aligned}k_{2i}^{-1}(\theta) &= \int_{d_{1i}}^{d_{2i}} \pi(\lambda|\theta) d\lambda \\ &= \sqrt{k} [\log(d_{2i}) - \log(d_{1i})]\end{aligned}$$

and

$$\begin{aligned}p_i(\lambda|\theta) &= k_{2i}(\theta) \pi(\lambda|\theta) \\ &= \frac{1}{\lambda [\log(d_{2i}) - \log(d_{1i})]}.\end{aligned}$$

In addition, the marginal reference prior for θ can be expressed as

$$\begin{aligned}\pi_i(\theta) &= \exp \left[\frac{1}{2} \int_{d_{1i}}^{d_{2i}} p_i(\lambda|\theta) \log \left(\frac{|\mathbf{I}|}{I_{11}} \right) d\lambda \right] \\ &\propto \frac{1}{\theta},\end{aligned}$$

from which the reference prior can be expressed as

$$\begin{aligned}\pi_{R2}(\theta, \lambda) &= \lim_{i \rightarrow \infty} \left[\frac{k_{2i}(\theta) \pi_i(\theta)}{k_{2i}(\theta_0) \pi_i(\theta_0)} \right] \pi(\lambda|\theta) \\ &\propto \frac{1}{\lambda \theta}\end{aligned}$$

for any fixed point θ_0 . This completes the proof. \square

Note that, since all the derived objective priors are improper, the corresponding posterior distribution should be proved to be proper. Since the Jeffreys prior $\pi_J(\lambda, \theta)$ and reference prior $\pi_{R2}(\theta, \lambda)$ have the same form, the notation $\pi_{JR}(\lambda, \theta)$ is used from now on.

2.2. Posterior Analysis

Let $\mathbf{x}_L = \{x_{L(1)}, \dots, x_{L(k)}\}$ be the observed lower record values. Then, the objective prior $\pi_{JR}(\lambda, \theta)$ results in the following marginal posterior density functions of λ and θ :

$$\begin{aligned}\pi_{JR}(\lambda|\mathbf{x}_L) &= \frac{\int_{\theta} L(\lambda, \theta) \pi_{JR}(\lambda, \theta) d\theta}{\int_{\theta} \int_{\lambda} L(\lambda, \theta) \pi_{JR}(\lambda, \theta) d\theta d\lambda} \\ &= \sum_{j=0}^{\infty} \frac{1}{\left(\sum_{i=1}^k x_{L(i)} + jx_{L(k)}\right)^k} \lambda^{k-1} \frac{(-\lambda)^j}{j!} \frac{\left(\sum_{i=1}^k x_{L(i)} - kx_{L(k)}\right)^k}{\Gamma(k)}\end{aligned}\quad (5)$$

and

$$\begin{aligned}\pi_{JR}(\theta|\mathbf{x}_L) &= \frac{\int_{\lambda} L(\lambda, \theta) \pi_{JR}(\lambda, \theta) d\lambda}{\int_{\theta} \int_{\lambda} L(\lambda, \theta) \pi_{JR}(\lambda, \theta) d\theta d\lambda} \\ &= \frac{\left(\sum_{i=1}^k x_{L(i)} - kx_{L(k)}\right)^k}{\Gamma(k)} \theta^{k-1} e^{-\theta(\sum_{i=1}^k x_{L(i)} - kx_{L(k)})},\end{aligned}\quad (6)$$

respectively. Note that the marginal posterior distribution of θ has a gamma distribution with the parameters k and $\sum_{i=1}^k x_{L(i)} - kx_{L(k)}$. Then, the Bayes estimators under the squared error loss function from the marginal posterior density functions (5) and (6) can be expressed as

$$\begin{aligned}\hat{\lambda}_{JR} &= \int_{\lambda} \lambda \pi_{JR}(\lambda|\mathbf{x}_L) d\lambda \\ &= k \left(1 - \frac{x_{L(k)}}{\sum_{i=1}^k x_{L(i)} - kx_{L(k)}}\right)^{-k}\end{aligned}$$

and

$$\hat{\theta}_{JR} = \frac{k}{\sum_{i=1}^k x_{L(i)} - kx_{L(k)}},$$

respectively.

In terms of σ , the corresponding marginal posterior density function can be expressed as

$$\pi_{JR}(\sigma|\mathbf{x}_L) = \frac{\left(\sum_{i=1}^k x_{L(i)} - kx_{L(k)}\right)^k}{\Gamma(k)} \sigma^{-k-1} e^{-\frac{1}{\sigma}(\sum_{i=1}^k x_{L(i)} - kx_{L(k)})}.$$

Since it is the PDF of an inverse gamma distribution with the parameters k and $\sum_{i=1}^k x_{L(i)} - kx_{L(k)}$, the Bayes estimator of σ under the squared error loss function can be expressed as

$$\hat{\sigma}_{JR} = \frac{\sum_{i=1}^k x_{L(i)} - kx_{L(k)}}{k-1}, \quad k > 1.$$

Theorem 3. From the Frequentist perspective, the estimator $\hat{\sigma}_{JR}$ is an unbiased estimator of σ .

Proof. According to Lemma 2 provided in Wang and Ye [12], independent and identically distributed random variables from the uniform distribution on $(0, 1)$ are defined as

$$U_j = \left(\frac{T_j}{T_{j+1}} \right)^j \\ = e^{j(X_{L(j+1)} - X_{L(j)})/\sigma}, \quad j = 1, \dots, k-1,$$

where

$$T_j = 2 \sum_{i=1}^j [\log F(x_{L(i-1)}) - \log F(x_{L(i)})] \\ = 2\lambda e^{-X_{L(j)}/\sigma}, \quad j = 1, \dots, k \quad (\log F(x_{L(0)}) \equiv 0)$$

are independent random variables having χ^2 distributions with $2j$ ($j = 1, \dots, k$) degrees of freedom. Then, the estimator $\hat{\sigma}_{JR}$ has a gamma distribution with the parameters $k-1$ and $(k-1)/\sigma$ for any $k > 1$ because

$$W(\sigma) = - \sum_{j=1}^{k-1} \log U_j \\ = \frac{\left(\sum_{i=1}^k X_{L(i)} - kX_{L(k)} \right)}{\sigma}$$

has a gamma distribution with the parameters $k-1$ and 1. This completes the proof. \square

The highest posterior density (HPD) credible intervals (CrIs) for λ and σ can be constructed by generating the MCMC samples from the marginal posterior density functions (5) and (6), respectively. However, since the marginal posterior density function (5) is not a well-known probability distribution, sampling from it is a difficult task. Instead, sampling can be indirectly performed from the relationship $\pi_{JR}(\lambda, \theta | \mathbf{x}_L) = \pi_{JR}(\lambda | \theta, \mathbf{x}_L) \pi_{JR}(\theta | \mathbf{x}_L)$ because the conditional posterior distribution $\pi_{JR}(\lambda | \theta, \mathbf{x}_L) \propto \lambda^{k-1} e^{-\lambda e^{-\theta x_{L(k)}}}$ has a gamma distribution with the parameters k and $e^{-\theta x_{L(k)}}$. To achieve this, θ should be generated from its marginal posterior distribution first, and then λ should be generated from its conditional posterior distribution given the generated value of θ . Finally, $\sigma = 1/\theta$ can be computed. Then, the $100(1-\alpha)\%$ equal-tails (ETs) and HPD CrIs can be constructed for $0 < \alpha < 1$ using the method proposed in Chen and Shao [13].

Under the prior $\pi_{R1}(\lambda, \theta)$, the resulting posterior $\pi_{R1}(\lambda, \theta | \mathbf{x}_L)$ is proper because $Q_2(\lambda) > 1$. However, since it is not possible to express it in a closed form as we know it, MCMC samples for λ and θ can be generated using the Metropolis–Hastings algorithm. For efficient mixing, the proposal variance-covariance structure is updated adaptively. For the corresponding Bayes estimators under the squared error loss function, the notations $\hat{\lambda}_{R1}$ and $\hat{\sigma}_{R1}$ are used.

2.3. Prediction

Let $X_{L(s)}$ ($s = k+1, k+2, \dots$) be the future lower record values. Since the sequence $\{X_{L(i)}, i = 1, 2, \dots\}$ is a Markov chain, the conditional density function of $X_{L(s)}$ given $\mathbf{X}_L = \mathbf{x}_L$ is the same as that of $X_{L(s)}$ given $X_{L(k)} = x_{L(k)}$. That is, it follows that

$$f_{X_{L(s)} | \mathbf{x}_L}(y) = f_{X_{L(s)} | x_{L(k)}}(y) \\ = \frac{[\log F(x_{L(k)}) - \log F(y)]^{s-k-1} f(y)}{\Gamma(s-k) F(x_{L(k)})}, \quad y < x_{L(k)} \quad (7)$$

by Ahsanullah [9]. Then, for the EGD with the CDF (1), (7) becomes

$$f_{X_{L(s)}|\mathbf{x}_L}(y) = \frac{\lambda \theta \left[\lambda \left(e^{-\theta y} - e^{-\theta x_{L(k)}} \right) \right]^{s-k-1}}{\Gamma(s-k)} e^{-\theta y - \lambda \left(e^{-\theta y} - e^{-\theta x_{L(k)}} \right)}, \quad -\infty < y < x_{L(k)} < \infty \quad (8)$$

and the corresponding Bayesian predictive density function can be expressed as

$$f_{X_{L(s)}|\mathbf{x}_L}^*(y) = \int_{\theta} \int_{\lambda} f_{X_{L(s)}|\mathbf{x}_L}(y; \lambda, \theta) \pi(\lambda, \theta | \mathbf{x}_L) d\lambda d\theta, \quad -\infty < y < x_{L(k)} < \infty,$$

where $\pi(\lambda, \theta | \mathbf{x}_L)$ is a general joint posterior distribution for (λ, θ) .

Note that it requires very complex and tedious computations. In fact, there is no guarantee that it can be expressed in a closed form. Instead, a much simpler approach is to use the pivotal quantity that can be obtained by the transformation of a random variable.

Let $H = \lambda \left(e^{-\theta y} - e^{-\theta x_{L(k)}} \right)$ in the conditional density function (8). Then, it has a gamma distribution with the parameters $s - k$ and 1 with a PDF of

$$f_H(h) = \frac{1}{\Gamma(s-k)} h^{s-k-1} e^{-h}, \quad h > 0$$

because $y < x_{L(k)}$ maps onto $h > 0$ and the Jacobian of the transformation is

$$\begin{aligned} J &= \frac{\partial}{\partial h} y \\ &= - \left[\lambda \theta \left(\frac{h}{\lambda} + e^{-\theta x_{L(k)}} \right) \right]^{-1}, \end{aligned}$$

which leads to the following algorithm for generating the MCMC samples $y_i (i = 1, \dots, N)$.

Step 1a. Generate h_i from the gamma distribution with the parameters $(s - k)$ and 1.

Step 1b. Generate λ_i and θ_i from the joint posterior distribution $\pi(\lambda, \theta | \mathbf{x})$.

Step 2. Compute

$$y_i = -\frac{1}{\theta_i} \log \left(\frac{h_i}{\lambda_i} + e^{-\theta_i x_{L(k)}} \right).$$

Step 3. Repeat steps 1 and 2, N times.

Then, the corresponding $100(1 - \alpha)\%$ predictive interval (PI) for $0 < \alpha < 1$ can be constructed using the method proposed in [13] as in the case of λ and θ . For the purpose of clarity, $X_{L(s)}^{JR} | x_{L(k)}$ and $X_{L(s)}^{R1} | x_{L(k)}$ are used to denote future lower record values under the priors $\pi_{JR}(\lambda, \theta)$ and $\pi_{R1}(\lambda, \theta)$, respectively.

3. Time Series Approach

Providing that record values are observed from time series of uncorrelated random variables sampled from continuous probability distributions, the proposed Bayesian method presented in the previous section is compared to the ARIMA and DLM time-series techniques as described below.

The ARIMA model is the most widely used approach to time-series forecasting. Conventionally, it is defined using three components (p, d, q) , where

- p denotes the order of the autoregressive (AR) term
- d denotes the number of differencing required to make the time series stationary
- q denotes the order of the moving average (MA) term.

Here, the autoregressive (AR) process assumes that the current value of the series y_t can be expressed as a function of p past values $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ in a form of

$$Y_t = \beta_0 + \sum_{j=1}^p \phi_j Y_{t-j} + \varepsilon_t$$

for $t \geq 1$, where $\beta_0 = (1 - \sum_{i=1}^p \phi_i) \mu$, μ is the mean of this process, $\phi_1, \phi_2, \dots, \phi_p$ are constants ($\phi_p \neq 0$), and ε_t is a weak white noise series with a mean of zero and a variance of σ_ε^2 . The MA process uses past forecast errors expressed as

$$Y_t = \mu + \sum_{j=1}^q \phi_j \varepsilon_{t-j} + \varepsilon_t,$$

that is, a weighted average of the past values of the white noise process ε_t . Then, the time series Y_t is an ARIMA(p, d, q) process if $\Delta^d Y_t$ is ARMA(p, q) obtained by combining the AR and MA terms, where Δ^d is the d th-order differencing operator. For non-stationary data, one usually fits an ARMA model after taking differences for the data until stationarity is achieved.

The second time-series approach considered in this study for comparison with the proposed method is the DLM. Let \mathbf{Y}_t be an m -dimensional vector observed at time t , while δ_t be a generally unobservable p -dimensional state vector of the system at time t . Then, the DLM can be defined as

$$\begin{cases} \mathbf{Y}_t = F_t \delta_t + v_t, & v_t \sim \mathcal{N}_m(0, V_t) \\ \delta_t = G_t \delta_{t-1} + w_t, & w_t \sim \mathcal{N}_p(0, W_t) \end{cases}$$

for each time $t \geq 1$ together with a prior distribution for the p -dimensional state vector at time $t = 0$, $\delta_0 \sim \mathcal{N}_p(m_0, C_0)$, where F_t and G_t are known matrices of $m \times p$ and $p \times p$, respectively, and V_t and W_t are variance matrices. Furthermore, it is assumed that the error sequences v_t and w_t are independent, and independent of δ_0 .

Note that the lower record value from a univariate time series has a strong trend of decreasing through time. Therefore, a DLM with

$$V_t = V, \quad F_t = F = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \delta_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}, \quad G_t = G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad W_t = W = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}$$

is considered, namely, the linear growth model (LGM)

$$\begin{cases} Y_t = \mu_t + v_t, & v_t \sim \mathcal{N}(0, V) \\ \mu_t = \mu_{t-1} + \beta_{t-1} + w_{t,1}, & w_{t,1} \sim \mathcal{N}(0, \sigma_\mu^2) \\ \beta_t = \beta_{t-1} + w_{t,2}, & w_{t,2} \sim \mathcal{N}(0, \sigma_\beta^2), \end{cases} \quad (9)$$

where μ_t and β_t denote the local level and local growth rate at time t , respectively, and $v_t, w_{t,1}$, and $w_{t,2}$ denote uncorrelated errors.

4. Sulfur Dioxide Data

This section demonstrates the superiority of the proposed Bayesian method by comparing it to the ARIMA and DLM methods.

The three methods are applied to the time-series data representing sulfur dioxide emissions in the United States (U.S.) from 1970 to 2017 (in 1000 tons) measured by the U.S. Environmental Protection Agency. Due to the implementation of the Acid Rain Program created under Title IV of the 1990 Clean Air Act, sulfur dioxide emissions have decreased significantly over the last decades through a cap and trade program for fossil-fuel powered plants. The observed volume of sulfur dioxide emissions and its descriptive statistics are presented in Figure 1 and Table 1, respectively. Note that each data point was divided by 1000 for computational convenience; given that the data continued to decrease during the

observation period as shown in Figure 1, they were all used without losing data during acquiring the lower record values.

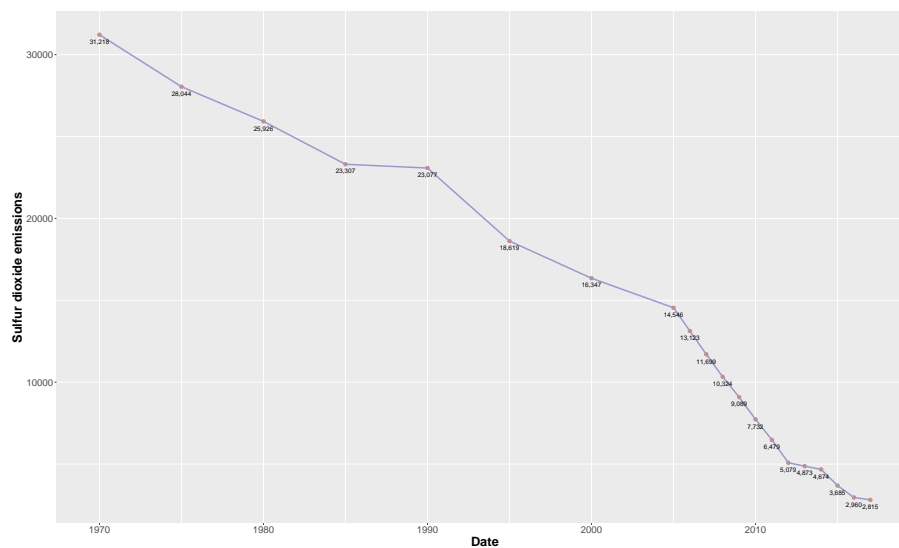


Figure 1. Observed sulfur dioxide lower record values.

Table 1. Descriptive statistics for the observed sulfur dioxide lower record values.

Minimum	Maximum	Mean	Median	Standard Deviation	Skewness	Kurtosis
2815	31,218	13.181	11.011	9.065	0.544	−1.147

To conduct the goodness-of-fit test for the observed sulfur dioxide data, the replicated data are first considered. If the estimated model is adequate, then it should look similar to the observed data. The replicated data are generated from the Bayesian predictive function

$$f_{X_{L(i)}}^*(x) = \int_{\lambda} \int_{\sigma} f_{X_{L(i)}}(x; \lambda, \theta) \pi(\lambda, \theta | \mathbf{x}_L) d\lambda d\theta$$

and denoted as $X_{L(i)}^{rep}$ for $i = 1, \dots, k$. Under each prior distribution, the correlation coefficient of the mean $E(X_{L(i)}^{rep})$ and observed lower record values r can then be computed. For further examination, the weighted mean squared error (WMSE)

$$WMSE = \frac{1}{k} \sum_{i=1}^k \frac{(x_{L(i)} - E(X_{L(i)} | \lambda, \sigma))^2}{Var(X_{L(i)} | \lambda, \sigma)}$$

is also computed. These results are reported in Figure 2.

It can be noticed from Figure 2 that the estimated models fit the observed sulfur dioxide lower record values very well, and the estimated models under the priors $\pi_{JR}(\lambda, \theta | \mathbf{x}_L)$ and $\pi_{R1}(\lambda, \theta | \mathbf{x}_L)$ provide almost the same results for the considered statistical criteria.

Table 2 reports the estimation results for the derived priors for comparison and corresponding maximum likelihood counterparts. For the maximum likelihood procedure, the maximum likelihood estimators (MLEs) $\hat{\lambda}$ and $\hat{\sigma}$ are obtained by maximizing the log-likelihood function (2), while the approximate $100(1 - \alpha)\%$ confidence intervals (CIs) are calculated based on the MLEs as

$$\hat{\lambda} \pm Z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\lambda})} \quad \text{and} \quad \hat{\sigma} \pm Z_{\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\theta})}{\theta^4}},$$

where Z_α denotes the upper α percentile point of the standard normal distribution, and $\widehat{Var}(\hat{\lambda})$ and $\widehat{Var}(\hat{\theta})$ are the diagonal elements of the asymptotic variance-covariance matrix of the MLE obtained by inverting the Fisher information matrix (3). For the shape parameter λ , the Bayes estimate $\hat{\lambda}_{R1}$ has a slightly lower value than the other estimates $\hat{\lambda}$ and $\hat{\lambda}_{JR}$ that have almost the same values. However, the 95% HPD CrI under the prior $\pi_{R1}(\lambda, \theta)$ has the shortest length. For the scale parameter σ , all estimates vary slightly, while the approximate 95% CI based on the $\hat{\sigma}$ has a shorter length than the other CrIs have.

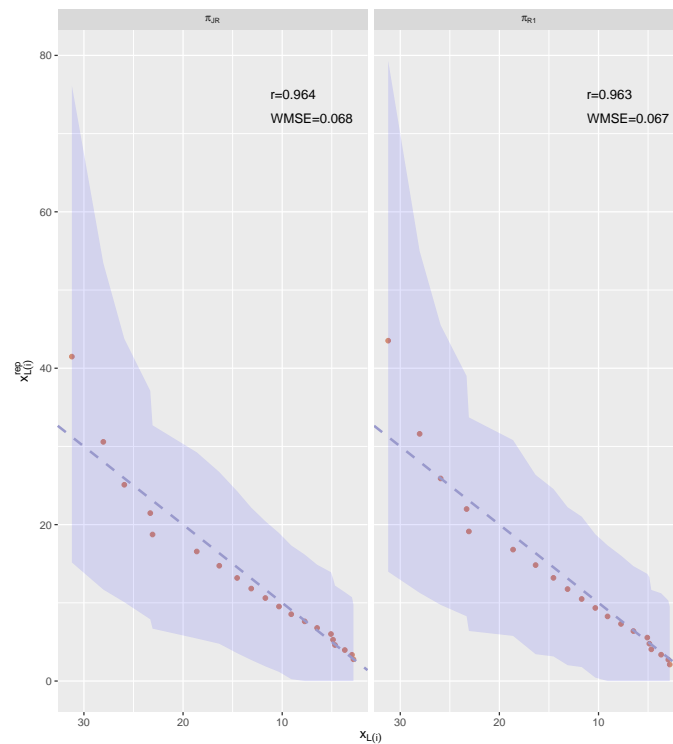


Figure 2. 95% regions for $X_{L(i)}^{rep}$ and scatter plots between $E(X_{L(i)}^{rep})$ and the observed lower record values.

Table 2. Results for the observed sulfur dioxide lower record values.

	$\hat{\lambda}$	$\hat{\lambda}_{JR}$	$\hat{\lambda}_{R1}$	$\hat{\sigma}$	$\hat{\sigma}_{JR}$	$\hat{\sigma}_{R1}$
Estimate	26.240	26.289	24.409	10.366	10.911	11.611
95% ETs	(14.411, 38.070)	(15.785, 39.653)	(14.874, 36.755)	(5.930, 14.801)	(6.991, 16.989)	(7.450, 18.242)
95% HPD	-	(15.082, 38.560)	(13.784, 35.388)	-	(6.493, 15.992)	(6.859, 17.039)

For prediction, the last lower record value is assumed to be known. As mentioned earlier, since no data are lost during acquiring the observed lower record values, the time-series analysis outlined in Section 3 can be conducted at the same time. Table 3 reports the prediction results for the next lower record value $X_{L(20)} | x_{L(19)}$. The R package **d1m** (Petrís [14]) was used to estimate the parameters and state vector in the LGM (9) with

$$\delta_0 \sim N_2 \left(\begin{bmatrix} 33.361 \\ -2.471 \end{bmatrix}, \begin{bmatrix} 0.593 & 0 \\ 0 & 0.190 \end{bmatrix} \right).$$

Table 3. Prediction results for the next lower record value.

	Mean	Median	95% PI		80% PI	
			ETs	HPD	ETs	HPD
$X_{L(20)}^{IR} x_{L(19)}$	2.359	2.560	(0.634, 2.945)	(1.115, 2.960)	(1.566, 2.899)	(2.009, 2.960)
$X_{L(20)}^{R1} x_{L(19)}$	2.275	2.507	(0.339, 2.945)	(0.905, 2.960)	(1.384, 2.894)	(1.870, 2.960)
$X_{L(20)}^{ARIMA}$	2.139	-	(0.082, 4.195)	-	(0.794, 3.483)	-
$X_{L(20)}^{DLM}$	2.357	-	(0.535, 4.179)	-	(1.165, 3.548)	-

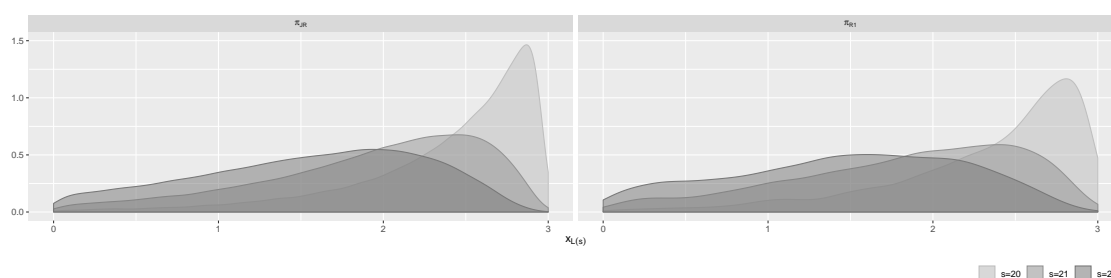
In addition, for the ARIMA model, ARIMA(0, 2, 1) was chosen as the best model in terms of the corrected Akaike Information Criterion (AICc), indicating that it has the smallest AICc value when $q = 1$ with $\phi = -0.705$ after differencing the data twice. The forecast accuracy are evaluated in terms of the mean absolute deviation (MAD), the mean square error (MSE), and the mean absolute percentage error (MAPE), defined respectively as

$$\begin{aligned} \text{MAD} &= \frac{1}{n} \sum_{t=1}^n |e_t|, \\ \text{MSE} &= \frac{1}{n} \sum_{t=1}^n e_t^2, \\ \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{x_t}, \end{aligned}$$

where $e_t = x_t - \hat{x}_t$ and \hat{x}_t is a point forecast of x_t . In this example, $n = k$ and $x_t = x_{L(t)}$. These results are reported in Table 4, which indicates that there is little difference in the predictive performance of the two models. Table 3 shows that the proposed Bayesian PI has a shorter length than those obtained for the considered time-series ARIMA model and DLM, especially under the prior $\pi_{JR}(\lambda, \theta)$. That is, the predictive result under the prior $\pi_{JR}(\lambda, \theta)$ shows the best performance in terms of uncertainty. Using the best performing Bayesian predictive model in terms of uncertainty, the Bayesian predictive density functions for the three future record values are estimated as the kernel density functions based on their MCMC samples. The results are plotted in Figure 3, which shows that both estimated Bayesian predictive models under the priors $\pi_{JR}(\lambda, \theta)$ and $\pi_{R1}(\lambda, \theta)$ have a greater variance as the future record time $L(s)$ increases.

Table 4. Forecast accuracy for the ARIMA model and DLM.

	MAD	MSE	MAPE
ARIMA(0,2,1)	0.649	0.927	0.062
DLM	0.664	0.817	0.062

**Figure 3.** Estimated kernel density for $X_{L(s)} | x_{L(19)}$ under the priors $\pi_{JR}(\lambda, \theta)$ and $\pi_{R1}(\lambda, \theta)$.

5. Conclusions

This paper defined the Jeffreys and reference priors for unknown parameters of the EGD based on record values and proposed a Bayesian method for predicting future record values. The method makes it very easy to generate MCMC samples for prediction. To validate the proposed method, it was compared to two time-series approaches, namely, the ARIMA model and DLM, using a sulfur dioxide emissions dataset. The results of the comparison demonstrated that the proposed method outperforms the time-series approaches in terms of uncertainty.

While there was no clear difference in the results of the goodness-of-fit tests among the proposed objective prior distributions when analyzing the observed data, the results of forecasting under the prior distribution $\pi_{JR}(\lambda, \theta)$ were better than those under the prior distribution $\pi_{R1}(\lambda, \theta)$; both derived prior distributions were proved to be valid.

Author Contributions: Y.K. and J.I.S. came up with the idea and developed the method presented in this paper. Both authors performed data analysis and wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. 2019R1I1A3A01062838).

Acknowledgments: We are grateful to the Editor-in-chief, Associate Editor, and the anonymous referees for their helpful comments.

Conflicts of Interest: There are no potential conflict of interest.

References

1. Chandler, K.N. The distribution and frequency of record values. *J. R. Stat. Soc. Ser. B* **1952**, *14*, 220–228. [CrossRef]
2. Coles, S.G.; Tawn, J.A. A Bayesian analysis of extreme rainfall data. *J. R. Stat. Soc. Ser. C* **1996**, *45*, 463–478. [CrossRef]
3. Madi, M.T.; Raqab, M.Z. Bayesian prediction of temperature records using the Pareto model. *Environmetrics* **2004**, *15*, 701–710. [CrossRef]
4. Wergen, G.; Hense, A.; Krug, J. Record occurrence and record values in daily and monthly temperatures. *Clim. Dyn.* **2013**, *42*, 1275–1289. [CrossRef]
5. Seo, J.I.; Kim, Y. Statistical inference on Gumbel distribution using record values. *J. Korean Stat. Soc.* **2016**, *45*, 342–357. [CrossRef]
6. Seo, J.I.; Kang, S.B. More efficient approaches to the exponentiated half-logistic distribution based on record values. *SpringerPlus* **2016**, *5*, 1433–1451. [CrossRef] [PubMed]
7. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting, and Control*, 3rd ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 1994.
8. West, M.; Harrison, J. *Bayesian Forecasting and Dynamic Models*, 2nd ed.; Springer: New York, NY, USA, 1997.
9. Ahsanullah, M. *Record Statistics*; Nova Science Publishers, Inc.: New York, NY, USA, 1995.
10. Jeffreys, H. *Theory of Probability and Inference*, 3rd ed.; Cambridge University Press: London, UK, 1961.
11. Beger, J.O.; Bernardo, J.M. Estimating a product of means: Bayesian analysis with reference priors. *J. Am. Stat. Assoc.* **1989**, *84*, 200–207 [CrossRef]
12. Wang, B.X.; Ye, Z.S. Inference on the Weibull distribution based on record values. *Comput. Stat. Data Anal.* **2015**, *83*, 26–36. [CrossRef]
13. Chen, M.H.; Shao, Q.M. Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comput. Graph. Stat.* **1999**, *8*, 69–92.
14. Petris, G. dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models. R Package Version 1.1-1. 2010. Available online: <http://CRAN.R-project.org/package=dlm> (accessed on 13 June 2020).

