

## Article

# Bayesian Forecasting of Dynamic Extreme Quantiles

Douglas E. Johnston

Farmingdale State College, The State University of New York, Farmingdale, NY 11735, USA;  
douglas.johnston@farmingdale.edu

**Abstract:** In this paper, we provide a novel Bayesian solution to forecasting extreme quantile thresholds that are dynamic in nature. This is an important problem in many fields of study including climatology, structural engineering, and finance. We utilize results from extreme value theory to provide the backdrop for developing a state-space model for the unknown parameters of the observed time-series. To solve for the requisite probability densities, we derive a Rao-Blackwellized particle filter and, most importantly, a computationally efficient, recursive solution. Using the filter, the predictive distribution of future observations, conditioned on the past data, is forecast at each time-step and used to compute extreme quantile levels. We illustrate the improvement in forecasting ability, versus traditional methods, using simulations and also apply our technique to financial market data.

**Keywords:** extreme value theory; quantile forecasting; particle filter; risk-management; value-at-risk



**Citation:** Johnston, D.E. Bayesian Forecasting of Dynamic Extreme Quantiles. *Forecasting* **2021**, *3*, 729–740. <https://doi.org/10.3390/forecast3040045>

Academic Editor: Sonia Leva

Received: 1 September 2021

Accepted: 7 October 2021

Published: 11 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Forecasting extreme quantiles, which is a part of extreme value (EV) analysis, is an important problem in many applications. For example, in analyzing ocean storm severity it is often critical to forecast an extreme quantile, such as the 99% quantile, for the maximum wave height over a certain time period to facilitate adequate construction design [1]. In financial risk-management, forecasting extreme quantiles, or so-called VaR analysis, is critical for establishing adequate bank capital levels [2]. In fact, extreme value analysis, and quantile forecasting, has found wide application in fields as varied as target detection [3], communication systems [4], image analysis [5], power systems [6], and population studies [7].

In this paper, which is an extension of the work presented in [8], we derive a recursive method for Bayesian forecasting of extreme quantiles where the underlying process is non-stationary. This differs from our previous work given our focus here is on the predictive distribution of future observations, rather than simply the model parameters, and we also properly invert the predictive distribution to obtain efficient quantile forecasts. In our previous work, we obtained improved parameter estimates, versus traditional methods, but our prediction ability, as measured by predictive  $p$ -values, was hampered by a number of issues. We made a number of improvements to our previous work including modifying the support assumptions on the observed data as well as implementing an adaptive particle filtering algorithm both of which resulted in demonstrable improvement in quantile forecasting.

We model the maximum of a block of data, which has an asymptotic general extreme value (GEV) distribution [9], via a non-linear state-space model where the parameters are driven by a hidden stochastic process with unknown static parameters. To recursively estimate the model parameters, we utilize a particle filter (PF) [10] and, in particular, we derive a Rao-Blackwellized particle filter (RBPF) [11] to marginalize the unknown, static parameters. Importantly, we derive a recursive solution for the predictive state density, which is required for the particle filter, and we design an algorithm that bears similarity to the well-known Kalman filter [12] and is therefore readily implemented.

The work presented here, and in [8], is an extension of some previous studies. In [13], a deterministic trend is applied to a subset of the GEV parameters, while in [14–16], a GEV parameter subset is dynamic with known state equation. More recently, Ref. [17] developed a dynamic model for the tail-index and a Gaussian-mixture approximation to linearize the estimation problem. Our work has extended these previous studies in a number of ways. First, under a reasonable assumption, we reduce the number of GEV parameters and model the remaining set as a vector Markov process, with unknown system and covariance matrix. Second, we recursively compute the marginalized state density, eliminating the need to estimate unwanted nuisance parameters and, in the process, we derive recursive expressions for the necessary sufficient statistics. This allows for a fast, real-time implementation without the need to batch-process observations. Lastly, we derive the predictive distribution, for the block-maximum, that we use to forecast extreme quantiles.

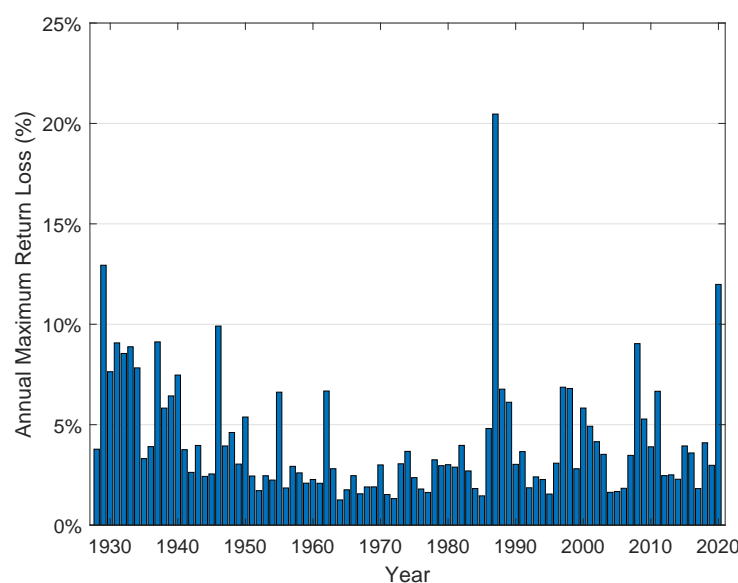
The paper is organized as follows. In the next section, we formulate the problem for the time-varying block-maximum and propose a Bayesian approach by deriving the recursive solution for marginalizing the nuisance parameters-Rao-Blackwellization. In Section 3, we implement our solution and derive our simplified likelihood function. We describe our algorithm in detail and show the method's ability to forecast extreme quantiles, outperforming traditional methods, using simulated data. In the last section, we discuss our results and use our approach to analyze S&P 500 stock market returns from 1928 to 2020. We also conclude the paper with ideas for further research.

## 2. Materials and Methods

The starting point for our analysis is a sequence of block-maximums denoted by  $y_k \in \mathbb{R}, k = 1, \dots, N$ . The  $k$ 'th block-maximum is the maximum of a strictly stationary time series,  $s_t$ , so that

$$y_k = \max_{N_{k-1} < t \leq N_k} s_t, \quad (1)$$

where the index  $t$  typically represents time and the  $k$ 'th block consists of data with time indices  $(N_{k-1}, N_k]$  and  $N_0 = 0$ . An example would be in financial markets where the underlying data is the daily return of a stock market index (e.g., S&P 500) and  $y_k$  is the largest daily return loss experienced in year  $k$  (see Figure 1). In this case,  $s_t = -r_t$ , where  $r_t = (p_t - p_{t-1})/p_{t-1}$  is the daily return defined as the percentage change in the price,  $p_t$ .



**Figure 1.** Annual block maximum of S&P 500 return loss (1928–2020).

Similar to the central-limit theorem, there is a limiting distribution for the normalized block-maximum of i.i.d. random variables (RVs). The Fisher-Tippett-Gnedenko (FTG)

theorem states that the only non-degenerate limiting distribution for the block-maximum, is in the generalized extreme value (GEV) family [18] with shape parameter, or tail-index,  $\xi$ . That is, the cumulative distribution function (cdf) of the block-maximum,  $y_k$ , properly normalized, converges to a distribution  $H_\xi$  as  $(N_k - N_{k-1}) \rightarrow \infty$  which has the Jenkinson-von Mises representation [19],

$$H_\xi(y) = \exp(-(1 + \xi y)^{-1/\xi}). \quad (2)$$

When the tail-index is positive ( $\xi > 0$ ), the resulting distribution is the Fréchet distribution with  $y \geq -1/\xi$ . The Fréchet is the limiting distribution for many heavy-tailed underlying random variables [20]. Upon normalization, a three parameter family is obtained and, therefore, we asymptotically model each of the block-maximum  $y_k \sim H_{\xi, \beta, \mu}(y_k) = H_\xi(\frac{y_k - \mu}{\beta})$ , with  $\beta > 0$  and support  $y \in [\mu - \beta/\xi, \infty)$ . We note that the FTG theorem's i.i.d. assumption can be relaxed to include most strictly stationary processes [21].

To capture the non-stationary effects and clustering often witnessed in real-world data (see Figure 1), we propose to model the parameters of the GEV distribution as time-varying and, to insure positivity for the shape and scale parameters, we define the unknown state vector as

$$\mathbf{x}_k = [\log(\xi_k), \log(\beta_k), \mu_k]^\top \quad (3)$$

and the state transition equation as

$$\mathbf{x}_k = \boldsymbol{\eta} + \boldsymbol{\Theta}\mathbf{x}_{k-1} + \mathbf{C}^{1/2}\mathbf{u}_k, \quad (4)$$

where  $\mathbf{x}_k \in \mathbb{R}^3$ ,  $k \in \mathbb{N}_0$ ,  $\mathbf{u}_k \sim \mathcal{N}(0, \mathbf{I}_3)$ ,  $\boldsymbol{\eta} \in \mathbb{R}^3$ , and  $\boldsymbol{\Theta}, \mathbf{C} \in \mathbb{R}^{3 \times 3}$ . We assume the matrices  $\boldsymbol{\eta}, \boldsymbol{\Theta}$ , and  $\mathbf{C}$  are unknown and the Gaussian process noise,  $\mathbf{u}_k$ , was chosen for its analytical tractability and maximal entropy property. Modeling the state as a first-order auto-regressive process, as in Equations (3) and (4), is similar to stochastic volatility models [22].

We specify the observation, or measurement, equation as

$$y_k = x_{3k} + e^{x_{2k}} w_k, \quad (5)$$

where  $e^{x_{2k}} = \beta_k$ ,  $x_{3k} = \mu_k$ , and  $w_k$  is distributed according to (2) and, hence, a function of  $e^{x_{1k}} = \xi_k$ . Specifically,  $y_k$  is a standard GEV RV that is scaled and translated and, combined, Equations (4) and (5) form a non-linear state-space model from which we wish to make Bayesian inferences.

The Bayesian inferences we wish to make come in the form of three probability density functions (pdf). The first is the state filtering density  $p(\mathbf{x}_k | y_{1:k})$  where  $y_{1:k} \equiv \{y_1, y_2, \dots, y_k\}$ . The second is the state predictive density  $p(\mathbf{x}_{k+1} | y_{1:k})$ . Lastly, and most germane to our goal, is the predictive distribution of the observation

$$p(y_{k+1} | y_{1:k}) = \int p(y_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | y_{1:k}) d\mathbf{x}_{k+1}. \quad (6)$$

Since our goal is to estimate extreme quantiles, we require the cdf of this predictive density,

$$F(y_{k+1} | y_{1:k}) = \int_{-\infty}^{y_{k+1}} p(\zeta | y_{1:k}) d\zeta, \quad (7)$$

which we can obtain under suitable conditions as

$$F(y_{k+1} | y_{1:k}) = E_{\mathbf{x}_{k+1} | y_{1:k}} [H_{\mathbf{x}_{k+1}}(y_{k+1})]. \quad (8)$$

In words, the cumulative predictive distribution is the expected GEV distribution conditioned on the state predictive density  $p(\mathbf{x}_{k+1} | y_{1:k})$  and our goal is to forecast  $\eta_\alpha = F^{-1}(\alpha | y_{1:k})$  or the  $\alpha\%$ -quantile.

To accomplish our goal, we need to derive a recursive solution for the filtering density,  $p(\mathbf{x}_k|y_{1:k})$ , and the state predictive density,  $p(\mathbf{x}_{k+1}|y_{1:k})$ . This can be done analytically for the case of known linear state-space models with Gaussian noise. However, with a non-linear state-space model, or non-Gaussian disturbances, a numerical approach is needed.

From standard Bayesian analysis, we can write

$$p(\mathbf{x}_{0:k}|y_{1:k}) \propto p(y_k|\mathbf{x}_k)p(\mathbf{x}_{0:k}|y_{1:k-1}), \quad (9)$$

where  $\mathbf{x}_{0:k} \equiv \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k\}$  is the *state stream* which are the state vectors from the initial state, at time 0, up to the current state, at time  $k$ . The likelihood function,  $p(y_k|\mathbf{x}_k)$ , highlights the assumption that the current observation depends only on the current value of the state vector. If we further assume that the state vector process is Markov, independent of the observations, we can write

$$p(\mathbf{x}_{0:k}|y_{1:k}) \propto \{p(y_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{0:k-1})\} p(\mathbf{x}_{0:k-1}|y_{1:k-1}), \quad (10)$$

which is the classic Bayesian recursion equation. If we can further invoke that the state process is first-order Markov, so that  $p(\mathbf{x}_k|\mathbf{x}_{0:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ , then we get the integral form of the Bayesian recursion [23],

$$p(\mathbf{x}_k|y_{1:k}) \propto p(y_k|\mathbf{x}_k) \int_{\mathbf{x}_{k-1}} p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|y_{1:k-1}) d\mathbf{x}_{k-1}, \quad (11)$$

which can be implemented with a standard particle filter [24].

In our case, where the static parameters in the state transition equation (4) are unknown, we can not directly implement Equations (10) or (11) since the current state depends on its complete history through the unknown parameters. The tact we take is to derive a Rao-Blackwellized particle filter via marginalizing  $p(\mathbf{x}_k, \boldsymbol{\eta}, \boldsymbol{\Theta}, \mathbf{C}|\mathbf{x}_{0:k-1})$ ,

$$\int_{\boldsymbol{\eta}, \boldsymbol{\Theta}, \mathbf{C}} p(\mathbf{x}_k|\boldsymbol{\eta}, \boldsymbol{\Theta}, \mathbf{C}, \mathbf{x}_{0:k-1}) dP(\boldsymbol{\eta}, \boldsymbol{\Theta}, \mathbf{C}|\mathbf{x}_{0:k-1}), \quad (12)$$

to obtain  $p(\mathbf{x}_k|\mathcal{T}_{k-1})$ , where  $\mathcal{T}_{k-1}$  are a set of sufficient statistics dependent only on the state stream up to time  $k-1$ . In doing so, we need an efficient, recursive solution to  $\mathcal{T}_{k-1}$  so that the filter can be implemented without the need for repeated, large matrix operations or the need to retain the complete state history.

To start, we note that Equation (4) can be written compactly as a general linear model for the complete state stream up to time  $k$  as

$$\mathbf{X}_k = \boldsymbol{\Phi} \mathbf{H}_k + \mathbf{U}_k, \quad (13)$$

where  $\mathbf{X}_k = [\mathbf{x}_k, \mathbf{x}_{k-1}, \dots, \mathbf{x}_1] \in \mathbb{R}^{p \times k}$ , with  $p$  being the number of state variables. The unknown matrix  $\boldsymbol{\Phi} = [\boldsymbol{\eta} \quad \boldsymbol{\Theta}] \in \mathbb{R}^{p \times p+1}$  and  $\mathbf{H}_k = [\mathbf{h}_k, \mathbf{h}_{k-1}, \dots, \mathbf{h}_1] \in \mathbb{R}^{q \times k}$  has columns given by  $\mathbf{h}_k^\top = [1 \quad \mathbf{x}_{k-1}^\top]$ . The noise term  $\mathbf{U}_k \in \mathbb{R}^{p \times k}$  is a random Normal matrix whose columns are i.i.d.  $\mathcal{N}(\mathbf{0}, \mathbf{C})$  with  $\mathbf{C}$  unknown. For the Rao-Blackwellized particle filter, the columns of  $\mathbf{X}_k$  are formed by the particles from 1 to  $k$  and the columns of  $\mathbf{H}_k$  include the past particles from 0 to  $k-1$ . Therefore, Ref. (13) is the state transition equation that describes the evolution of the state stream and, in the filter, each particle stream operates under its own version of this multivariate regression model.

The predictive density for the general linear model, with a non-informative, Jeffreys prior [25], results in a multivariate Student-t distribution [26], which for a  $p$ -dimensional random vector is denoted as  $t_p(v, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto [1 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})/v]^{-(v+p)/2}$  with  $v$  degrees of freedom, mean  $\boldsymbol{\mu}$ , and scaling matrix  $\boldsymbol{\Sigma}$ . Using the non-informative prior,  $p(\boldsymbol{\Theta}, \boldsymbol{\Theta}, \mathbf{C}^{-1}) \propto |\mathbf{C}|^{(p+1)/2}$ , we can write the marginal predictive density for  $\mathbf{x}_{k+1}$  as a multivariate Student-t distribution with  $v_k = k - 2p$  degrees of freedom, i.e.,

$$p(\mathbf{x}_{k+1}|\mathbf{X}_k, \mathbf{H}_k, \mathbf{h}_{k+1}) \sim t_p(v_k, \hat{\mathbf{x}}_{k+1}, \hat{\boldsymbol{\Sigma}}_{k+1}).$$

The mean of this distribution is derived as

$$\hat{\mathbf{x}}_{k+1} = \mathbf{S}_{XH}^k (\mathbf{S}_{HH}^k)^{-1} \mathbf{h}_{k+1} = \hat{\Phi}_k \mathbf{h}_{k+1} \quad (14)$$

and the scale matrix is

$$\hat{\Sigma}_{k+1} = \frac{1 + \mathbf{h}_{k+1}^\top (\mathbf{S}_{HH}^k)^{-1} \mathbf{h}_{k+1}}{v_k} (\mathbf{S}_{XX}^k - \hat{\Phi}_k \mathbf{S}_{XH}^{k\top}), \quad (15)$$

with the matrices defined as

$$\mathbf{S}_{XX}^k = \mathbf{X}_k \mathbf{X}_k^\top, \quad \mathbf{S}_{XH}^k = \mathbf{X}_k \mathbf{H}_k^\top, \quad \mathbf{S}_{HH}^k = \mathbf{H}_k \mathbf{H}_k^\top. \quad (16)$$

As is, Equations (14)–(16) can be used to compute the state predictive density and to extrapolate particles from time  $k$  to time  $k + 1$ . That said, the resulting matrix operations are expensive computationally and a recursive solution is desired. Letting  $\mathbf{P}_k = (\mathbf{S}_{HH}^k)^{-1}$ , we derived recursive expressions for the mean and the scale matrix as follows:

$$\mathbf{K}_k = \mathbf{h}_k^\top \mathbf{P}_{k-1} [1 + \mathbf{h}_k^\top \mathbf{P}_{k-1} \mathbf{h}_k]^{-1}, \quad (17)$$

$$\mathbf{P}_k = \mathbf{P}_{k-1} [\mathbf{I} - \mathbf{h}_k \mathbf{K}_k], \quad (18)$$

$$\hat{\Phi}_k = \hat{\Phi}_{k-1} + (\mathbf{x}_k - \hat{\mathbf{x}}_k) \mathbf{K}_k, \quad (19)$$

$$\begin{aligned} \hat{\Sigma}_{k+1} &= \frac{v_k - 1}{v_k} \left\{ \frac{1 + \mathbf{h}_{k+1}^\top \mathbf{P}_k \mathbf{h}_{k+1}}{1 + \mathbf{h}_k^\top \mathbf{P}_{k-1} \mathbf{h}_k} \right\} \\ &\quad \times \left\{ \hat{\Sigma}_k + (\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^\top / (v_k - 1) \right\}, \end{aligned} \quad (20)$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\Phi}_k \mathbf{h}_{k+1}. \quad (21)$$

These equations bear a resemblance to the recursive equations found in the Kalman filter [27] and are readily implemented with minimum computational and memory requirements. The specific algorithm we use is discussed in the next section and we note that the above procedure can likewise be used in the case of a linear observation equation with unknown observation matrix.

### 3. Results

To implement our solution, we first make a simplifying assumption that reduces our state vector to  $\mathbf{x}_k = [\log(\xi_k), \log(\beta_k)]^\top$ . Recall that for the GEV distribution, with  $\xi > 0$ , the support for the observation is  $y \in [\mu - \beta/\xi, \infty)$ , where we have dropped the subscript  $k$  for simplicity. To eliminate the parameter  $\mu$ , we constrain  $H_{\xi, \beta, \mu}(y_0) = \alpha_0$  allowing us to solve for  $\mu = -H_{\xi, \beta, y_0}^{-1}(\alpha_0)$ . This is a generalization of the constraint used in [8] where the support was  $y \in [0, \infty)$  or  $\alpha_0 = y_0 = 0$  resulting in  $\mu = \beta/\xi$ . The main issue in the previous study was that small values for the parameter  $\xi$  resulted in a substantial portion of the left tail with negligible probability since  $H_{\xi, \beta, \beta/\xi}^{-1}(\alpha) \rightarrow \infty$  as  $\xi \rightarrow 0^+ \forall \alpha > 0$  resulting in sub-par prediction performance. While the parameter estimates in [8] were quite reasonable, and outperformed the maximum likelihood (ML) method, quantile estimates were biased.

Under our support constraint, we can write

$$\mu = y_0 - \frac{\beta}{\xi} [(-\ln(\alpha_0))^{-\xi} - 1] \quad (22)$$

and the now two-parameter GEV distribution is

$$H_{\xi, \beta}(y) = \exp \left[ - \left( \frac{\xi}{\beta} (y - y_0) + (-\ln(\alpha_0))^{-\xi} \right)^{-1/\xi} \right] \quad (23)$$

with support  $y \in [y_0 - \frac{\beta}{\xi}(-\ln(\alpha_0))^{-\xi}, \infty)$ . The likelihood function,  $p(y|\mathbf{x})$ , can then be written as

$$p(y|\mathbf{x}) = \frac{d}{dy} H_{\xi, \beta}(y) = \frac{1}{\beta} H_{\xi, \beta}(y) \left( \frac{\xi}{\beta} (y - y_0) + (-\ln(\alpha_0))^{-\xi} \right)^{-(1+\xi^{-1})}, \quad (24)$$

where the substitutions  $\xi = e^{x_1}$  and  $\beta = e^{x_2}$  can readily be made.

Armed with the likelihood in Equation (24) and the sufficient statistics, computed recursively in Equations (17)–(21), we are ready to implement our version of the RBPF (see Algorithm 1).

---

**Algorithm 1** RBPF recursive implementation.

---

Initialization ( $k = 0, m = 1 : M$  particles):

- Generate  $\tilde{p} > 2p + 1$  random particles,  $\mathbf{x}_{0:\tilde{p}}^{(m)}$  from (13) with random  $\Phi, \mathbf{C}$ .
- Compute  $\mathbf{S}_{XX}^{\tilde{p}}, \mathbf{S}_{XH}^{\tilde{p}}$ , and  $\mathbf{S}_{HH}^{\tilde{p}}$  from (16).
- Initialize  $\hat{\Phi}_0 = \mathbf{S}_{XH}^{\tilde{p}}(\mathbf{S}_{HH}^{\tilde{p}})^{-1}$ ,  $\hat{\mathbf{x}}_1$  from (14) and  $\hat{\Sigma}_1$  from (15).
- Let  $\mathbf{P}_0 = (\mathbf{S}_{HH}^{\tilde{p}})^{-1}$  and  $v_k = \tilde{p} - 2p$ .
- Set particle weights  $w^{(m)} = 1/M$  and  $M_1 = M$ .

Recursive Loop ( $k = 1 : N$ ):

- Sample  $\mathbf{x}_k^{(m)-} \sim t_p(v_k, \hat{\mathbf{x}}_k, \hat{\Sigma}_k)$  for  $m = 1, \dots, M_k$ . (particle extrapolation)
  - Compute predictive statistics:  $\hat{\mathbf{x}}_k^- = \sum_{m=1}^M w^{(m)} \mathbf{x}_k^{(m)-}$
  - Compute quantile estimates,  $p$ -values, and  $M_k$  (see Algorithm 2)
  - Systematically resample particles,  $\mathbf{x}_k^{(m)-} \rightarrow \mathbf{x}_k^{(m)+}$ , using  $p(y_k|\mathbf{x}_k)$  in (24)
  - Compute contemporaneous statistics:  $\hat{\mathbf{x}}_k^+ = \sum_{m=1}^M w^{(m)} \mathbf{x}_k^{(m)+}$
  - Compute  $\mathbf{K}_k$  from (17). Update  $\mathbf{P}_{k-1} \rightarrow \mathbf{P}_k$  via (18) and  $\hat{\Phi}_{k-1} \rightarrow \hat{\Phi}_k$  via (19)
  - Update to  $\hat{\Sigma}_{k+1}$  from (20) and  $\hat{\mathbf{x}}_{k+1}$  from (21)
  - $k = k + 1, v_k = v_k + 1$ , and repeat loop
- 

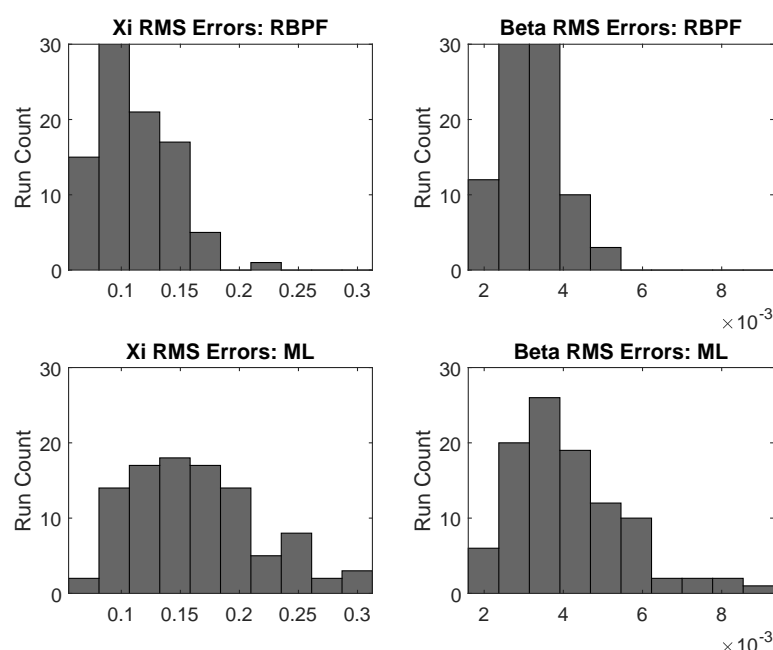
In Algorithm 1, estimates for  $\xi_k$  and  $\beta_k$  are produced by a weighted sum of the particles  $\mathbf{x}_k^{(m)+}$  available after processing the observation,  $y_k$ , at time  $k$ . To illustrate the performance of the algorithm, we simulated the stochastic GEV parameters from an uncoupled, mean-reverting, AR(1) process as follows

$$\log \xi_k = (1 - 0.95) \times \log(0.45) + 0.95 \log \xi_{k-1} + 0.1 u_{1,k}, \quad (25)$$

$$\log \beta_k = (1 - 0.9) \times \log(0.012) + 0.9 \log \beta_{k-1} + 0.1 u_{2,k}, \quad (26)$$

where  $u_{1,k}$  and  $u_{2,k}$  are independent  $\mathcal{N}(0, 1)$ . The process was chosen with  $\bar{\xi} = 0.45$  and  $\bar{\beta} = 1.25\%$  to be representative of time-series encountered in heavy-tailed phenomena such as financial markets and we refer the reader to [8] for representative simulation examples.

We produced 225 samples for each simulation run and we initialized the filter with  $M_1 = 1024$  particles. At each point in time we computed estimates for the GEV parameters  $\hat{\xi}_k$  and  $\hat{\beta}_k$ . We also computed the maximum likelihood (ML) estimates for the GEV parameters from (24) and used at least 25 observations ( $k \geq 25$ ) before reporting estimation errors. For each of the 100 simulation runs, we computed a root-mean-square (RMS) error for the parameter estimates, across time, and shown in Figure 2 are the histograms for the RMS errors for both the RBPF and ML method. We can see the RBPF method outperforms with an average RMS error for  $\hat{\xi}$  of 0.11 vs. 0.17 for the ML method. Similarly, the average RMS error for  $\hat{\beta}$  was 0.32% vs. 0.41% for the ML method.



**Figure 2.** Histogram of RMS errors for RBPF vs ML for 100 simulation runs.

One of the improvements we made to our original algorithm was to use systematic resampling of the existing set of particles versus standard multinomial resampling which improves the filter’s performance. The other improvement was to adapt the number of particles based on an assessment of the filter’s predictive performance as in [28] although we modified their method so that new particles, if needed, are generated from the initial prior distribution. Finally, the quantile estimates,  $\hat{\eta}_\alpha = \hat{F}^{-1}(\alpha|y_{1:k})$  for  $\alpha \in [0, 1]$ , based on averaging  $H_{\mathbf{x}_{k+1}}(\cdot)$  over the particles that approximate the predictive state distribution, were improved. In particular, we computed an estimate of the predictive distribution, via approximating Equation (8), versus using  $E_{\mathbf{x}_{k+1}|y_{1:k}}[H_{\mathbf{x}_{k+1}}^{-1}(\alpha)]$ . Algorithm 2 provides the details of the method used for quantile estimation, computing  $p$ -values, and to adapt the number of particles in the filter.

To test our approach, we examined the predictive performance of our method, outlined in Algorithm 2, versus the maximum likelihood. We applied a chi-square test as described in [29]. At each time  $k$ , we effectively partitioned the support of the predictive cdf estimate,  $\hat{F}(y_{k+1}|y_{1:k})$  into  $B$  equiprobable buckets. For the ML method, we did the same using the ML estimates of the GEV parameters as “truth”. We then compared the  $(k + 1)$ st observation,  $y_{k+1}$ , to each bucket interval to find its location and to create an empirical frequency distribution. If the model is true, this frequency distribution will be a sample estimate of a uniform distribution and the statistic  $D$  is approximately chi-square with  $B - 1$  degrees of freedom (See Algorithm 2 for details).

For our test, we chose  $B = 20$  intervals and shown in Figure 3 are the frequency distributions for the entire 100 runs. We note that, in total, there were 20,000 sets of quantiles forecasted and each forecast was based solely on past data. More importantly, each prediction was based on a varied amount of past data. For example, in each simulation run, the first set of quantile forecasts were based on just 25 observations ( $y_{1:25}$ ) and the last forecasts were based all that run’s observations except the last ( $y_{1:224}$ ). We see that the RBPF method has a more uniform frequency distribution compared to the ML. The ML, similar to our previous work in [8], tended to under-estimate the 5% quantile, leading to more observations than expected. More importantly, the extreme quantiles were overestimated leading to less observations exceeding those thresholds. While one might argue that the bias to higher thresholds, for those extreme quantiles, is at least being “conservative”, it is clearly more beneficial to having a more accurate cdf and quantile forecast, particularly

since these quantile forecasts are typically used to compute exceedance probabilities over multi-block periods.

---

**Algorithm 2** Compute quantile estimates,  $p$ -values, and  $M_k$ .

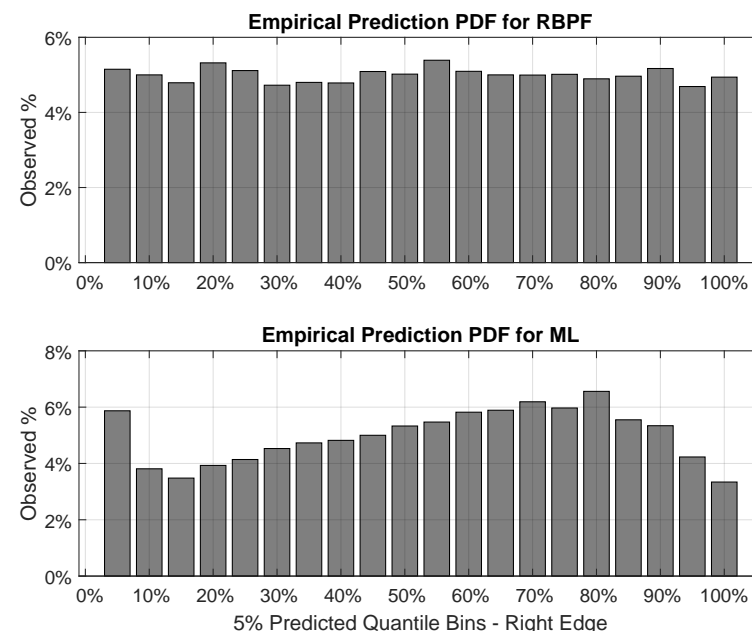
---

Initialization (prior to Recursive Loop)

- Set  $\alpha_j = j/B$  for  $j = 0, \dots, B$  ( $B = \#$  buckets)
- Create  $y$ -grid:  $\tilde{y} = 0 : y_{step} : y_{max}$  and set  $[\hat{\xi}_k^-, \hat{\beta}_k^-]^T = \exp(\hat{x}_k^-)$ .
- Initialize counter  $C(j) = 0$  for  $j = 1, \dots, B$

Within Recursive Loop ( $k \geq N_{pred}$ )

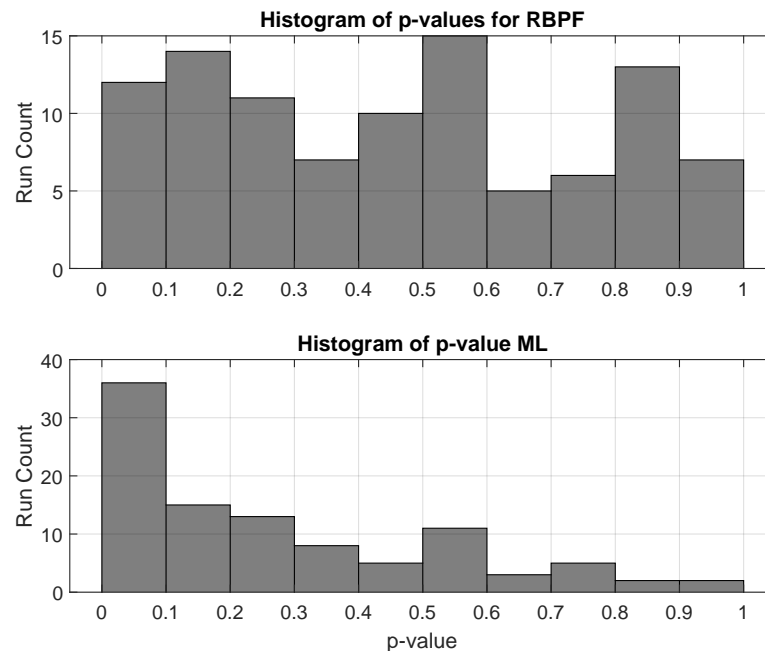
- Compute predictive CDF:  $\hat{F}(\tilde{y}|y_{1:k}) = \sum_{m=1}^M w^{(m)} H_{\hat{\xi}_k^-, \hat{\beta}_k^-}(\tilde{y})$ .
  - Invert predictive CDF:  $\eta_{\alpha_j} = \hat{F}^{-1}(\alpha_j) \forall j$
  - Find index  $j^*$  such that  $y_k \in [\eta_{\alpha_{j^*-1}}, \eta_{\alpha_{j^*}})$  for  $j^* \in \{1, \dots, B\}$
  - Increment counter  $C(j^*)$
  - If  $k \geq 2 * N_{pred}$  (compute  $p$ -values)
    - $N_c = \sum_{j=1}^B C(j)$
    - Compute  $\chi^2$  statistic:  $D = B(N_c - 1) \sum_{j=1}^B (C(j)/N_c - 1/B)^2$
    - Compute  $p$ -value using  $\chi_{B-1}^2(D)$  test
    - If  $p$ -value  $< p_{low}$  (double # of particles)
      - \*  $M_{new} = M_k - \min(2M_k, M_{max})$
      - \* Initialize  $M_{new}$  new particles (See Initialization in Algorithm 1)
      - \*  $M_k = M_k + M_{new}$
    - If  $p$ -value  $> p_{high}$  (halve # of particles)
      - \* Randomly discard  $M_k - \max(M_k/2, M_{min})$  existing particles
      - \*  $M_k = \max(M_k/2, M_{min})$
- 



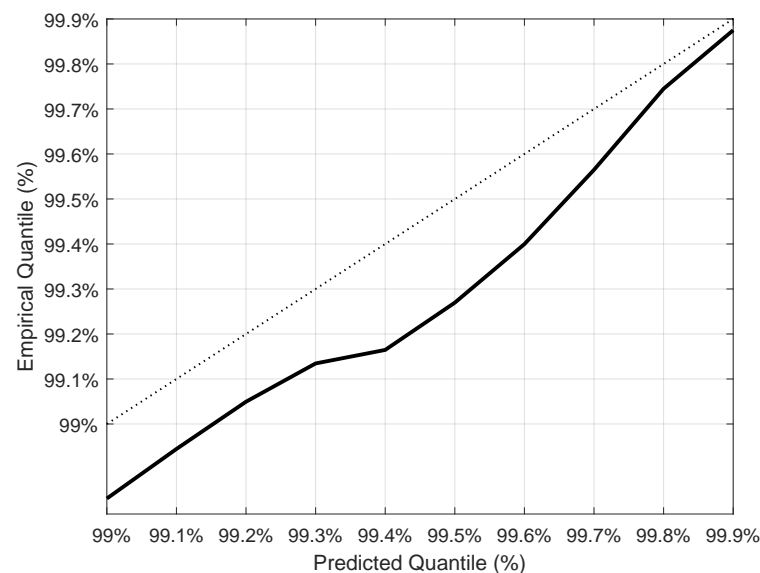
**Figure 3.** Empirical frequency distribution of observation predictions for RBPF vs ML simulation.

To further illustrate the improvement in the performance of our approach, we show the histogram of the final  $p$ -values, for the 100 simulation runs, in Figure 4. The ML method has about 35 of the  $p$ -values in the range of  $[0,0.1)$ , compared to an expected amount of 10, and most of the  $p$ -values are less than 0.5. This indicates that one would reject the ML method as an alternative hypothesis in most instances. In our previous work, our

simulated  $p$ -values were worse than the ML method due to the reasons cited previously. Our new RBPF method, derived herein, has  $p$ -values close to the expected number for each bucket (i.e., 10), and produces quite reasonable quantile forecasts. To further see the accuracy of forecasting extreme quantiles with the RBPF, we plotted the empirical quantile estimate versus the predicted in Figure 5.



**Figure 4.** Distribution of final  $p$ -values for RBPF vs ML simulation.



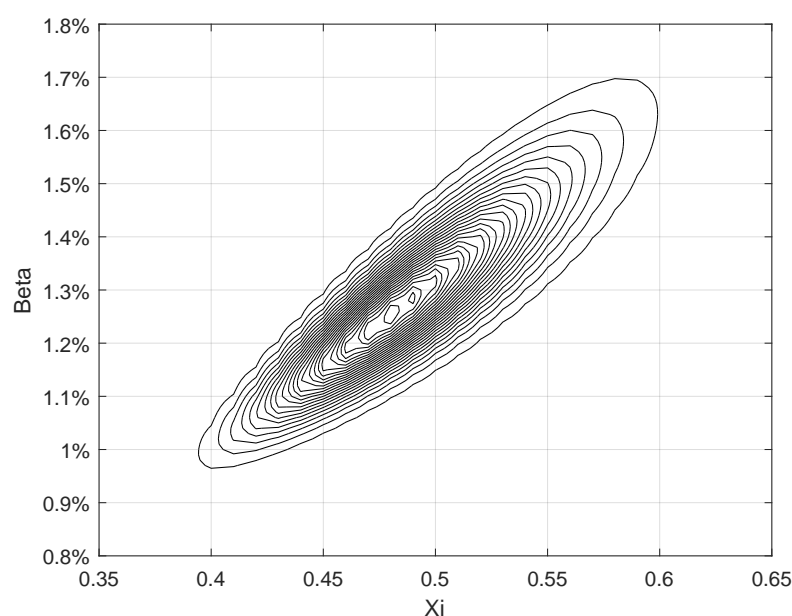
**Figure 5.** Empirical extreme quantile estimates versus predicted for the RBPF.

As seen in Figure 5, at the predicted 99.9%-quantile, 99.87% of the observations were below the threshold. Stated differently, 26 out of the 20,000 predictions exceeded the forecasted 99.9%-quantile threshold, which is six more than expected. We find this to be excellent given that many of the forecasts are made with limited data in hand. At the 99%-quantile, 0.15% additional observations exceeded our forecasts. Out of the 20,000 predictions, 200 exceedances were expected versus the 230 were observed. Our simulation results bear out the view that our methodology not only outperforms traditional quantile forecast methods (e.g., ML) but also our previous work in [8].

#### 4. Discussion

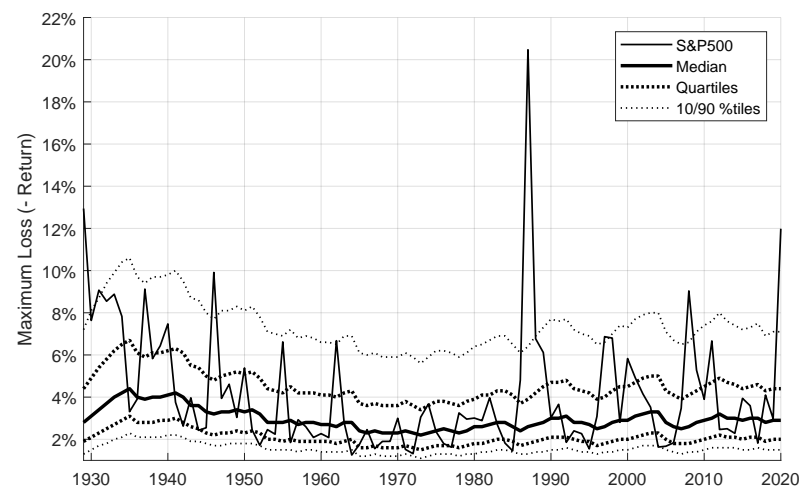
From our results, illustrated in the previous section, we see that our methodology offers improvements in forecasting extreme quantiles, particularly when the time-series is non-stationary. This is true in many applications, where extreme behaviour tends to cluster in time, such as in finance, or trend over time, such as in climatology. To illustrate the application of our approach to real-world data, we applied it the data shown in Figure 1 which are the block-maximums of daily losses, or negative returns expressed in percent, for the S&P 500 stock market index, a widely used proxy for the overall stock market.

For completeness, we show in Figure 6 the likelihood surface for all of the data from 1928–2020 (93 years/block-maximums). While the ML estimates, using all of the data, are  $\hat{\xi}_{ML} = 0.48$  and  $\hat{\beta}_{ML} = 1.26\%$ , one can see there is a large degree of uncertainty in the GEV parameters. For example, there is considerable probability that  $\xi > 0.5$ , which would imply an underlying time-series with infinite variance and this has implications for financial models that rely on finite second moments (e.g., Black-Scholes). The Bayesian approach is to use all of the information available from the data to produce posterior densities from which to produce parameter estimates and, more importantly, to forecast. This is particularly important when forecasting quantiles since they can be quite sensitive to point estimates given the highly non-linear form of the GEV cdf.



**Figure 6.** Likelihood surface for GEV parameters using S&P 500 block-maximums.

Shown in Figure 7 are the S&P 500 block-maximums along with quantile forecasts over time. For the 90% quantile forecast, 8 out of 92 observed stock-market returns exceeded the forecast, or 8.7% of the time. This is well within what one would expect. The current, forecasted 90% quantile is 7.1% and the 99% quantile forecast is 23.8%. We should expect to see this type of market "crash," or worse, about once a century. While the range of the forecasted 90% quantile, from 5.6% in the 1970s to 8% post financial crisis, may seem minimal, we do note that more extreme quantile estimates have significant variation. For example, the 99% quantile forecast went from 18% to 25% and the 99.5% quantile went from 25.8% to 35% over the same period. This would be of use to a financial institution when constructing risk scenarios for stress testing their portfolios. Last year, in 2020, the block-maximum was a 12% loss, during the covid pandemic crisis, which was at the 96% quantile. We should expect to see at least this 12% loss about once every 25 years and it should not be considered unusual highlighting the importance of quantile forecasting—bad things do happen!



**Figure 7.** S&P 500 max annual loss and predicted quantiles.

It is worth pointing out that the  $p$ -value based on the forecasted quantiles was a robust 0.75 versus 0.01 for the ML method. While one must be wary in using  $p$ -values to accept a model, we feel that the results we have shown in this paper clearly indicate an improved method of forecasting. But much work needs to be done. For one, we simply used block-maximums in this paper, which one can argue wastes data. We would like to extend our approach further to use more higher-order statistics, or exceedances above a threshold. We would also like to broaden our simulation study and apply our results to other data sets, particularly in the area of climatology. Lastly, additional computational efficiencies should be explored to allow real-time implementation to high-frequency data.

**Funding:** This research was funded in part by a Farmingdale State College, Mathematics Department, Summer Research Grant.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Daily closing prices for the S&P 500 index were downloaded from Yahoo! finance, <https://finance.yahoo.com>, and are available upon request, access date: 1 September 2021.

**Acknowledgments:** The author thanks Farmingdale State College and Carlos Marques for their continued support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

cdf	Cumulative density function
EV	Extreme Value
FTG	Fisher-Tippet-Gnedenko
GEV	Generalized extreme value
ML	Maximum likelihood
pdf	Probability density function
PF	Particle filter
RBPF	Rao-Blackwellized particle filter
RMS	Root mean square
RV	Random variable
S&P	Standard and Poor
VaR	Value at Risk

## References

1. Northrop, P.J.; Attalides, N.; Jonathan, P. Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *J. R. Stat. Soc. Appl. Stat.* **2017**, *66*, 93–120. [\[CrossRef\]](#)
2. Johnston, D.E.; Djurić, P.M. The science behind risk management: A signal processing perspective. *Signal Process. Mag.* **2011**, *28*, 26–36. [\[CrossRef\]](#)
3. Broadwater, J.B.; Chellappa, R. Adaptive threshold estimation via extreme value theory. *IEEE Trans. Signal Process.* **2010**, *58*, 490–500. [\[CrossRef\]](#)
4. Resnick, S.I.; Rootzeń, H. Self-similar communication models and very heavy tails. *Ann. Appl. Probab.* **2000**, *10*, 753–778. [\[CrossRef\]](#)
5. Roberts, S.J. Extreme value statistics for novelty detection in biomedical data processing. *IEEE Proc. Sci. Meas. Technol.* **2000**, *147*, 363–367. [\[CrossRef\]](#)
6. Shenoy, S.; Gorinevsky, D. Estimating long tail models for risk trends. *IEEE Signal Process. Lett.* **2015**, *22*, 968–972. [\[CrossRef\]](#)
7. Anderson, S.C.; Branch, T.A.; Cooper, A.B.; Dulvy, N.K. Black-swan events in animal populations. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3252–3257. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Johnston, D.E.; Djurić, P.M. A recursive Bayesian model for extreme values. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019.
9. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2004.
10. Arulampalam, M.S.; Maskell, N.G.; Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **2002**, *50*, 174–188. [\[CrossRef\]](#)
11. Chen, R.; Wang, X.; Liu, J.S. Adaptive joint detection and decoding in flat fading channels via mixture Kalman filtering. *IEEE Trans. Inf. Theory* **2000**, *46*, 2079–2094. [\[CrossRef\]](#)
12. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [\[CrossRef\]](#)
13. Hundedcha, Y.; St-Hilaire, A.; Quarda, T.; Adlouni, S.E. A nonstationary extreme value analysis for the assessment of changes in extreme annual wind speed over the Gulf of St. Lawrence, Canada. *Environmetrics* **2013**, *24*, 51–62. [\[CrossRef\]](#)
14. Huerta, G.; Sansó, B. Time-varying models for extreme values. *Environ. Ecol. Stat.* **2007**, *14*, 285–299. [\[CrossRef\]](#)
15. Toulemonde, G.; Guillou, A.; Naveau, P. Particle filtering for Gumbel-distributed daily maxima of methane and nitrous oxide. *J. Appl. Meteorol. Climatol.* **2008**, *47*, 2745–2759. [\[CrossRef\]](#)
16. Wei, Y.; Huerta, G. Dynamic generalized extreme value modeling via particle filters. *Communications Stat.* **2017**, *46*, 6324–6341. [\[CrossRef\]](#)
17. Mao, G.; Zhang, Z. Stochastic tail index model for high frequency financial data with Bayesian analysis. *J. Econom.* **2018**, *205*, 470–487. [\[CrossRef\]](#)
18. Embrechts, P.; Kluppelberg, C.; Mikosch, T. *Modelling Extremal Events*; Springer: New York, NY, USA, 2003.
19. Smith, R.L. Statistics of extremes with applications in environment, insurance and finance. In *Extreme Values in Finance, Telecommunications and the Environment*; Finkenstädt, B.; Rootzeń, H., Eds.; Chapman and Hall CRC: Boca Raton, FL, USA, 2003; pp. 1–78.
20. Resnick, S. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*; Springer: New York, NY, USA, 2008.
21. McNeil, A.; Frey, R.; Embrechts, P. *Quantitative Risk Management*; Princeton University Press: Princeton, NJ, USA, 2005.
22. Jacquier, E.; Polson, N.G.; Rossi, P.E. Bayesian analysis of stochastic volatility models. *J. Bus. Econ. Stat.* **1994**, *12*, 371–417.
23. Kramer, S.C.; Sorenson, H.W. Bayesian parameter estimation. *IEEE Trans. Autom. Control.* **1988**, *33*, 217–222. [\[CrossRef\]](#)
24. Djurić, P.M.; Bugallo, M.F. Particle filtering. In *Adaptive Signal Processing*; Adali, T.; Haykin, S., Eds.; Wiley & Sons: Hoboken, NJ, USA, 2010; pp. 271–331.
25. Jaynes, E.T. Prior Probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241. [\[CrossRef\]](#)
26. Geisser, S. Bayesian estimation in multivariate analysis. *Ann. Am. Stat.* **1965**, *36*, 150–159. [\[CrossRef\]](#)
27. Meinhold, R.J.; Singpurwalla, N.D. Understanding the Kalman filter. *Am. Stat.* **1983**, *37*, 123–127.
28. Elvira, V.; Míguez, J.; Djurić, P.M. Adapting the number of particles in sequential Monte Carlo methods through an online scheme for convergence assessment. *IEEE Trans. Signal Process.* **2017**, *65*, 1781–1794. [\[CrossRef\]](#)
29. Djurić, P.M.; Khan, M.; Johnston, D.E. Particle filtering of stochastic volatility modeled with leverage. *IEEE J. Sel. Top. Signal Process.* **2012**, *6*, 327–336. [\[CrossRef\]](#)