Computational Statistics and Data Analysis 56 (2012) 3241–3259

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Generalized extreme value distribution with time-dependence using the AR and MA models in state space form

Jouchi Nakajima[a], Tsuyoshi Kunihama[a], Yasuhiro Omori[b,*], Sylvia Frühwirth-Schnatter[c]

[a] Department of Statistical Science, Duke University, Durham 27708-0251, USA
[b] Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan
[c] Department of Applied Statistics, Johannes Kepler University Linz, Altenbergerstr. 69, 4040 Linz, Austria

ARTICLE INFO

Article history:
Available online 29 April 2011

Keywords:
Extreme values
Generalized extreme value distribution
Markov chain Monte Carlo
Mixture sampler
State space model
Stock returns

ABSTRACT

A new state space approach is proposed to model the time-dependence in an extreme value process. The generalized extreme value distribution is extended to incorporate the time-dependence using a state space representation where the state variables either follow an autoregressive (AR) process or a moving average (MA) process with innovations arising from a Gumbel distribution. Using a Bayesian approach, an efficient algorithm is proposed to implement Markov chain Monte Carlo method where we exploit an accurate approximation of the Gumbel distribution by a ten-component mixture of normal distributions. The methodology is illustrated using extreme returns of daily stock data. The model is fitted to a monthly series of minimum returns and the empirical results support strong evidence of time-dependence among the observed minimum returns.

1. Introduction

Extreme value theory has been applied in various fields from environmental sciences to financial econometrics. The salient feature of the extreme value analysis is to assess the extremal behavior of random variables. Previous studies often focused on independently and identically distributed random variables to consider the statistical property of their maxima or minima using parametric models (see, e.g., Leadbetter et al., 2004; Coles, 2001).

In recent decades, dynamic extreme value models have attracted considerable attention in the literature to investigate time-dependence or structural change of extremes. The extension to time series of extreme values can be accomplished by assuming time-dependence for the underlying stochastic state of the extreme value process. A conventional approach to capture time-dependence is to consider an autoregressive process for the model parameters of the extreme value distribution using a state space representation. An earlier example is Smith and Miller (1986), and several extensions have been explored (Gaetan and Grigoletto, 2004; Huerta and Sansó, 2007).

Another approach is to consider the class of max-stable processes (see e.g., Smith, 2003). The moving maxima process, for instance, is defined as the maximum of the past latent Fréchet innovations multiplied by weights summing to one. It is a stationary stochastic process with the marginal distribution equal to the Fréchet distribution. These processes have been extended to the maxima of moving maxima process (Deheuvels, 1983; Kunihama et al., in press) and the multivariate maxima of moving maxima process (e.g., Smith and Weissman, 1996, Zhang and Smith, 2004 and Chamú Morales, 2005). Smith (2003) and Kunihama et al. (in press) provide some applications of these classes to the financial extreme data.

* Corresponding author. Tel.: +81 3 5841 5516; fax: +81 3 5841 5521.
E-mail addresses: jouchi.nakajima@stat.duke.edu (J. Nakajima), tsuyoshi.kunihama@stat.duke.edu (T. Kunihama), omori@e.u-tokyo.ac.jp (Y. Omori), sylvia.fruehwirth-Schnatter@jku.at (S. Frühwirth-Schnatter).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.04.017

In this paper, we address a novel estimation methodology for an extreme value model with time-dependence which is induced by a time-varying latent variable of a non-Gaussian state space model. We begin with the generalized extreme value (GEV) distribution given by

$$\Pr(Y_t \leq y_t) = \exp\left\{-\left(1 + \xi \frac{y_t - \mu}{\psi}\right)^{-\frac{1}{\xi}}\right\}, \tag{1}$$

where $\psi > 0$, and $1 + \xi(y_t - \mu)/\psi > 0$. This GEV distribution is commonly used for the analysis of maxima or minima of some larger set of random variables. The subset of the GEV family with $\xi = 0$, which is interpreted as the limit of (1) as $\xi \to 0$, is known as the Gumbel or the Type I extreme value distribution. Further, in the case of $\mu = 0$ and $\psi = 1$, the distribution of the Gumbel random variable, defined by α_t, is obtained by

$$G(x) \equiv \Pr(\alpha_t \leq x) = \exp(-\exp(-x)). \tag{2}$$

The mean and variance of α_t are $E(\alpha_t) = c_0$, and $\mathrm{Var}(\alpha_t) = \pi^2/6$ ($\equiv c_1$), respectively, where c_0 is the Euler constant.

Suppose that Y_t follows the GEV distribution of (1), and define

$$\alpha_t \equiv \log\left\{\left(1 + \xi \frac{Y_t - \mu}{\psi}\right)^{\frac{1}{\xi}}\right\}. \tag{3}$$

Then, α_t follows the Gumbel distribution $G(\cdot)$ of (2), which yields the following relation between Y_t and α_t,

$$Y_t = \mu + \psi \frac{\exp(\xi\alpha_t) - 1}{\xi}, \tag{4}$$

where $\alpha_t \sim G$. Based on this relation, the current paper proposes the dynamic extreme value model where the measurement equation is (4) with an additional idiosyncratic shock (as in Chamú Morales (2005)). The time-dependence is introduced either in form of an autoregressive (AR) process or a moving average (MA) process for the latent variable α_t. We assume that the innovation of the process follows the Gumbel distribution. These measurement system and underlying process form the nonlinear and non-Gaussian state space model.

To the best of our knowledge, this is the first attempt to discuss such a time-dependence in literature. The conventional approach for the time-varying GEV model is to assume that the parameters of the GEV distribution follow the AR process with normal-distributed errors (e.g., Gaetan and Grigoletto, 2004; Huerta and Sansó, 2007). However, as shown above, there exists a natural way to incorporate time-dependence into the underlying latent process such that the response of the observation equation follows the GEV distribution. The key of this modeling is that the innovations of the latent process follow the Gumbel distribution. As a side remark, the second term in (4) can be regarded as the Box–Cox transformation of $\exp(\alpha_t)$, a common technique for non-Gaussian modeling.

The current work responds to the increasingly pressing need to address the time-dependence for the extreme value process. In principle, a theoretical-limit model of the extreme value process is applicable only to some restricted cases under the presence of temporal dependence. However, in application, Coles (2001) suggests that "it is usual to adopt a pragmatic approach of using the standard extreme models as basic templates that can be enhanced by statistical modeling". One justification to motivate the proposed model in this paper is that its time-dependence is an approximation to explain the time-varying structure of extreme values underlying the time series of interest. If the time-dependence is estimated to exist, the time-dependence in the proposed dynamic model would provide a better approximation to the underlying process than a static extreme value model. Also, it would presumably contribute to prediction.

Because the proposed time-dependent GEV model takes the form of the nonlinear and non-Gaussian state space model, it is computationally intensive to implement maximum likelihood estimation. We instead employ a Bayesian estimation approach using the Markov chain Monte Carlo (MCMC) method (see, e.g., Chib, 2001, Koop, 2003, Geweke, 2005 and Gamerman and Lopes, 2006). To facilitate the MCMC estimation method, we exploit an accurate approximation of the Gumbel density based on the ten-component mixture of normal distributions provided by Frühwirth-Schnatter and Frühwirth (2007). Introducing a mixture indicator variable reduces the original model to a conditionally Gaussian nonlinear state space model that allows an efficient sampling of the states as in Omori and Watanabe (2008). This approach, called a *mixture sampler*, is inspired by the related literature of Kim et al. (1998) and Omori et al. (2007) in the context of stochastic volatility models, where the log χ_1^2 density is approximated by a mixture of normal distributions.

The rest of paper is organized as follows. Section 2 defines the GEV model where the state variables follow an AR(1) process, and develops the MCMC estimation algorithm. An efficient particle filter is also proposed to compute the likelihood function. Section 3 introduces the GEV model where the state variables follow an MA(1) process. Section 4 illustrates the proposed model and estimation method using simulated data. In Section 5, the application to extreme returns of daily stock data is provided with the discussion of posterior estimates and posterior predictive analysis as well as the model comparisons based on marginal likelihood and forecasting performance. Finally, Section 6 concludes.

2. The GEV-AR model

2.1. Model specification

Let $y = \{y_1, \dots, y_n\}$ be a sequence of extreme values. We define the GEV model with a first order AR process, which we label GEV-AR model, by

$$y_t = \mu + \psi \frac{\exp(\xi\alpha_t) - 1}{\xi} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2),\ t = 1, \dots, n, \tag{5}$$

$$\alpha_{t+1} = \phi\alpha_t + \eta_t, \quad \eta_t \sim G,\ t = 1, \dots, n-1, \tag{6}$$

where $|\phi| < 1$. The state variable α_t is assumed to follow the stationary AR(1) process driven by the innovations following the Gumbel distribution defined in (6). Furthermore, we introduce in (5) a measurement error ε_t which is assumed to follow a normal distribution.

Allowing ϕ to be different from zero introduces dependence over time. The distribution of the time series y_t is driven by the time-varying state variable α_t, which is the weighted sum of the current innovation η_{t-1} and past innovations η_{t-j-1} for $j = 1, \dots, t-2$, weighted by ϕ^j; meanwhile, the innovations arise from the Gumbel rather than the normal distribution. Note that if ϕ and σ^2 both were zero, then $\{y_t\}_{t=1}^n$ would be a sequence of independently and identically distributed observations from the GEV distribution defined in (4).

For estimation purposes, we need to specify the distribution of the initial state α_1. Ideally, the distribution of α_1 would be the stationary distribution of the process (6). Evidently, the mean and the variance of this distribution are given by $c_0/(1-\phi)$ and $c_1/(1-\phi^2)$, respectively, where c_0 and c_1 are the mean and the variance of the Gumbel distribution. However, because it is not possible to work out the whole distribution, we approximate the distribution of α_1 by a normal distribution with the same mean and variance as the stationary distribution, *i.e.*,

$$\alpha_1 \sim N(c_0/(1-\phi), c_1/(1-\phi^2)). \tag{7}$$

2.2. Bayesian estimation

The unknown model parameters of the GEV-AR model are equal to $\omega \equiv (\lambda, \sigma^2, \phi)$, where $\lambda = (\mu, \psi, \xi)'$. For estimation we pursue a Bayesian approach based on assuming prior independence between λ, σ^2 and ϕ, *i.e*, $\pi(\lambda, \sigma^2, \phi) = \pi(\lambda)\pi(\sigma^2)\pi(\phi)$. Concerning σ^2, we use the conditionally conjugate prior $\sigma^2 \sim \mathrm{IG}(n_0/2, S_0/2)$, where IG denotes the inverse gamma distribution. No such conditionally conjugate priors exist for λ and ϕ. Our subsequent analysis allows complete flexibility concerning the choice of $\pi(\lambda)$ and $\pi(\phi)$. In our case studies we will assume prior independence among all components of ω, with μ and ξ following a normal, ψ following a Gamma and $(\phi+1)/2$ following a beta prior distribution.

For practical Bayesian estimation, we use MCMC methods to sample from the posterior distribution (see, e.g., Chib, 2001, Koop, 2003, Geweke, 2005 and Gamerman and Lopes, 2006 for a recent review of this technique). As common for state space models, we employ data augmentation by introducing the latent state process $\alpha = \{\alpha_t\}_{t=1}^n$ as missing data. Sampling of the latent state process presents a challenge, because the state equation is non-Gaussian, while the observation equation is nonlinear. In the present paper, we use the idea of auxiliary mixture sampling and approximate the nonlinear, non-Gaussian state space models (5) and (6) by an accurate finite mixture of nonlinear Gaussian state space models. This allows to sample the state variables from their posterior distribution efficiently through the MCMC algorithm.

2.2.1. Auxiliary mixture sampler

The idea of auxiliary mixture sampling has been well developed in the context of stochastic volatility models by approximating the log χ_1^2 density by a finite normal mixture (Kim et al., 1998; Omori et al., 2007). The mixture normal densities whose parameters do not depend on other parameters make the MCMC estimation highly efficient for non-Gaussian state space models. Recently, this idea has been extended to efficient estimation of non-Gaussian latent variables models like random-effects and state space models for binary, categorical, multinomial, and count data by approximating the density of the Gumbel distribution by a finite normal mixture (Frühwirth-Schnatter and Frühwirth, 2007; Frühwirth-Schnatter and Wagner, 2006; Frühwirth-Schnatter et al., 2009).

Following this work, we approximate the exact probability density function $g(\eta_t)$ of the Gumbel distribution underlying the innovations η_t in state equation (6) by a normal mixture of K components:

$$g(\eta_t) = \exp(-\eta_t - \mathrm{e}^{-\eta_t}) \approx \hat{g}(\eta_t) = \sum_{i=1}^{K} p_i f_N(\eta_t | m_i, v_i^2), \tag{8}$$

where $f_N(\eta_t|m_i, v_i^2)$ denotes the probability density function of a normal distribution with mean m_i and variance v_i^2. Frühwirth-Schnatter and Frühwirth (2007) propose an accurate mixture approximation based on $K = 10$ components where the selection of (p_i, m_i, v_i^2) for $i = 1, \dots, 10$ is reproduced in Table 1.

Table 1
Selection of (p_i, m_i, v_i^2) by Frühwirth-Schnatter and Frühwirth (2007).

i	p_i	m_i	v_i^2
1	0.00397	5.09	4.5
2	0.0396	3.29	2.02
3	0.168	1.82	1.1
4	0.147	1.24	0.422
5	0.125	0.764	0.198
6	0.101	0.391	0.107
7	0.104	0.0431	0.0778
8	0.116	−0.306	0.0766
9	0.107	−0.673	0.0947
10	0.088	−1.06	0.146

As a second step of data augmentation we introduce a mixture indicator variable, $s_t \in \{1, \ldots, K\}$ for $t = 1, \ldots, n-1$. Conditional on $s \equiv \{s_1, \ldots, s_{n-1}\}$, Eqs. (5) and (6) form a nonlinear Gaussian state space model, where

$$y_t = \mu + \psi \frac{\exp(\xi\alpha_t) - 1}{\xi} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \; t = 1, \ldots, n, \tag{9}$$

$$\alpha_{t+1} = m_{s_t} + \phi\alpha_t + v_{s_t} u_t, \quad u_t \sim N(0, 1), \; t = 1, \ldots, n-1, \tag{10}$$

and that α_1 follows the normal distribution defined in (7).

We implement the following algorithm to sample from the joint posterior density $\pi(\omega, s, \alpha|y)$.

Algorithm 1 (*MCMC Algorithm for the GEV-AR Model*)**.**

1. Generate $(\mu, \psi, \xi) \mid \sigma^2, \alpha, y$.
2. Generate $\sigma^2 \mid \mu, \psi, \xi, \alpha, y$.
3. Generate $\phi \mid \alpha$.
4. Generate $s \mid \phi, \alpha$.
5. Generate $\alpha \mid \omega, s, y$.

Note that in this scheme all model parameters ω are sampled without conditioning on s. In particular, the conditional posterior distribution of ϕ is marginalized over s which is expected to make sampling more efficient. Sampling from the inverse Gamma posterior $\sigma^2 \mid \mu, \psi, \xi, \alpha, y$ is straightforward, while sampling from $(\mu, \psi, \xi) \mid \sigma^2, \alpha, y$ and $\phi|\alpha$ requires the implementation of a Metropolis–Hastings step. To obtain high acceptance rates, we use proposal densities based on the mode and the Hessian of the conditional posterior densities. Details are provided in Appendix A.1.

Sampling the latent mixture indicator variables s is a standard step in finite mixture modeling (see e.g. Frühwirth-Schnatter, 2006). To sample the latent state process α, we apply the block sampler developed by Omori and Watanabe (2008) for nonlinear Gaussian state space models. Such blocking is known to produce more efficient draws than a single-move sampler which samples one state α_t at a time given the others states α_s $(s \neq t)$ (Shephard and Pitt, 1997). Within each block a Metropolis–Hastings step is employed based on normal proposal densities obtained from a Taylor expansion of the nonlinear mean appearing in the observation equation (5). Again, details are provided in Appendix A.1.

2.2.2. *Reweighting to correct for the mixture approximation error*

The MCMC draws of ω and α obtained by Algorithm 1 are not draws from the exact posterior distribution $\pi(\omega, \alpha|y)$, but draws from an approximate distribution $\hat{\pi}(\omega, \alpha|y)$ which is the marginal posterior of the approximate model (9)–(10) where the exact transition density $f(\alpha_{t+1}|\alpha_t, \phi) = g(\alpha_{t+1} - \phi\alpha_t)$ is substituted by the approximate density $\hat{f}(\alpha_{t+1}|\alpha_t, \phi) = \hat{g}(\alpha_{t+1} - \phi\alpha_t)$ given by (8). Though the normal mixture distribution (8) provides a good approximation to the Gumbel distribution, this subsection describes how to correct for the minor approximation error.

Let ω^j and α^j denote the jth sample from the approximated model, for $j = 1, \ldots, M$, where M is the number of iteration. To obtain a sample from the exact posterior distribution $\pi(\omega, \alpha|y)$, we resample the draws from the approximate posterior density with weights proportional to

$$w_j = \frac{w_j^*}{\sum_{i=1}^{M} w_i^*}, \qquad w_j^* = \frac{\pi(\omega^j, \alpha^j|y)}{\hat{\pi}(\omega^j, \alpha^j|y)} = \frac{f(\alpha^j|\phi^j)}{\hat{f}(\alpha^j|\phi^j)}, \quad j = 1, \ldots, M, \tag{11}$$

where $f(\alpha|\phi) = f(\alpha_1|\phi)\prod_{t=1}^{n-1} f(\alpha_{t+1}|\alpha_t, \phi)$ is given by the product of the exact transition densities, while $\hat{f}(\alpha|\phi) = f(\alpha_1|\phi)\prod_{t=1}^{n-1}\hat{f}(\alpha_{t+1}|\alpha_t, \phi)$ is given by the product of the approximate transition densities, *i.e.*, $\hat{f}(\alpha_{t+1}|\alpha_t, \phi) = \sum_{s_t=1}^{K} p_{s_t} f_N(\alpha_{t+1}|\phi\alpha_t + m_{s_t}, v_{s_t}^2)$. The posterior moments are obtained by computing the weighted average of the MCMC draws (Kim et al., 1998).

2.3. A new efficient particle filter

In addition to MCMC estimation, we propose a new efficient particle filter method to compute the likelihood function $f(y|\omega)$ for a fixed model parameter ω. This allows to perform model comparison and to compute goodness-of-fit statistics for model diagnostics (see Section 5).

The basic idea is to sample from a target posterior distribution recursively with the help of an importance function that well approximates the target density. For the GEV-AR model, using the measurement density $f(y_t|\alpha_t, \omega)$ from (5) and the evolution density $f(\alpha_{t+1}|\alpha_t, \omega)$ from (6), the associated particle filter is based on

$$f(\alpha_{t+1}, \alpha_t|Y_{t+1}, \omega) \propto f(y_{t+1}|\alpha_{t+1}, \omega)f(\alpha_{t+1}|\alpha_t, \omega)f(\alpha_t|Y_t, \omega),$$

where $Y_t = \{y_j\}_{j=1}^t$. The particles are drawn from $f(\alpha_t|Y_t, \omega)$ to yield a discrete uniform approximation $\hat{f}(\alpha_t|Y_t, \omega)$ to $f(\alpha_t|Y_t, \omega)$.

The simple particle filter (PF) uses $f(\alpha_{t+1}|\alpha_t, \omega)$ as an importance function, but it is known to produce inefficient estimates of the likelihood. Alternatively, the auxiliary particle filter (APF, Pitt and Shephard, 1999) is often used as an efficient filter in various fields. In the analysis of extreme values, it is pointed out (e.g., Chamú Morales, 2005) that such a filter often generates particles with almost zero importance weights for the extreme observations. This is because the APF constructs an importance function by exploiting the mean or the mode of the prior distribution of the state α_{t+1} given α_t. Many particles with zero weights result in a poor approximation of the filtering density and in inaccurate estimates of the likelihood, as we shall see in Section 5.5. To overcome this difficulty, we propose a simple but efficient particle filter where we base the importance function directly on the observation y_{t+1} to approximate the target density well even when there are extreme values.

First, we take the expectation of the error term in the observation equation, and consider the approximation of $y_{t+1} \approx \mu + \psi\{\exp(\xi\alpha_{t+1}) - 1\}/\xi$. Define m_{t+1} to replace α_{t+1} as

$$m_{t+1} \equiv \frac{1}{\xi}\log\left(1 + \xi\frac{y_{t+1} - \mu}{\psi}\right)_+ \approx \alpha_{t+1}, \quad t = 1, \ldots, n-1, \tag{12}$$

where $y_+ = \max(y, 0)$. Because m_{t+1} can be considered as the most likely value of the state α_{t+1} given y_{t+1}, we use the importance function

$$g(\alpha_{t+1}, \alpha_t^i|Y_{t+1}, \omega) = g(\alpha_{t+1}|y_{t+1}, \omega)\hat{f}(\alpha_t^i|Y_t, \omega),$$
$$g(\alpha_{t+1}|y_{t+1}, \omega) = \exp\{-(\alpha_{t+1} - m_{t+1})\}\exp\{-e^{-(\alpha_{t+1} - m_{t+1})}\}.$$

Note that this importance density generates α_{t+1} from the Gumbel distribution with the mode m_{t+1}. We propose the following particle filter:

1. Initialize $t = 1$, and generate $\alpha_1^i \sim N(c_0/(1-\phi), c_1/(1-\phi^2))$, for $i = 1, \ldots, I$.
   (a) Compute $w_1^i = f(y_1|\alpha_1^i)$ and $W_1^i = F(y_1|\alpha_1^i)$, where F denotes the distribution function of y_t given α_t, and save $\bar{w}_1 = \frac{1}{I}\sum_{i=1}^I w_1^i$, and $\bar{W}_1 = \frac{1}{I}\sum_{i=1}^I W_1^i$.
   (b) Set $\hat{f}(\alpha_1^i|y_1, \omega) = w_1^i/\sum_{j=1}^I w_1^j,\ i = 1, \ldots, I$.
2. Generate $(\alpha_{t+1}^i, \alpha_t^i)$, $i = 1, \ldots, I$, from the importance function $g(\alpha_{t+1}, \alpha_t|Y_{t+1}, \omega)$.
   (a) Compute
   $$w_t^i = \frac{f(y_{t+1}|\alpha_{t+1}^i, \omega)f(\alpha_{t+1}^i|\alpha_t^i, \omega)\hat{f}(\alpha_t^i|Y_t, \omega)}{g(\alpha_{t+1}^i, \alpha_t^i|Y_{t+1}, \omega)} = \frac{f(y_{t+1}|\alpha_{t+1}^i, \omega)\hat{f}(\alpha_{t+1}^i|\alpha_t^i, \omega)}{g(\alpha_{t+1}^i|y_{t+1}, \omega)},$$
   $$W_t^i = \frac{F(y_{t+1}|\alpha_{t+1}^i, \omega)f(\alpha_{t+1}^i|\alpha_t^i, \omega)}{g(\alpha_{t+1}^i|y_{t+1}, \omega)}, \quad i = 1, \ldots, I,$$
   and save $\bar{w}_t = \frac{1}{I}\sum_{i=1}^I w_t^i$, and $\bar{W}_t = \frac{1}{I}\sum_{i=1}^I W_t^i$.
   (b) Set $\hat{f}(\alpha_{t+1}^i|Y_{t+1}, \omega) = w_t^i/\sum_{j=1}^I w_t^j,\ i = 1, \ldots, I$.
3. Increase t by 1, and go to 2.

It can be shown that as $I \to \infty$, $\bar{w}_{t+1} \xrightarrow{p} f(y_{t+1}|Y_t, \omega)$ and $\bar{W}_{t+1} \xrightarrow{p} F(y_{t+1}|Y_t, \omega)$, then it follows that $\sum_{t=1}^n \log\bar{w}_t \xrightarrow{p} \sum_{t=1}^n \log f(y_t|Y_{t-1}, \omega)$.

2.4. Extension to a threshold model

The GEV distribution is commonly applied to the extremes that exceed a high threshold (see e.g., Coles, 2001). We show an extension of the time-dependent GEV model utilized to the threshold model. Let δ denote the threshold, and $\tilde{y}_t$ be a censored variable. We assume that the extreme value, y_t, is observed only when $\tilde{y}_t \geq \delta$. In the GEV-AR model, we modify the observation equation in the following way:

$$y_t = \begin{cases} \tilde{y}_t & (\text{if } \tilde{y}_t \geq \delta), \\ \text{N.A.} & (\text{if } \tilde{y}_t < \delta), \end{cases}$$

$$\tilde{y}_t = \mu + \psi \frac{\exp(\xi\alpha_t) - 1}{\xi} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2),\ t = 1, \ldots, n.$$

The extension of the MCMC algorithm discussed in Section 2.2 to this threshold model is straightforward. We only need to sample the auxiliary variable $\tilde{y}_t$ for t such that $y_t < \delta$. The conditional posterior distribution of $\tilde{y}_t$ reads:

$$\tilde{y}_t \mid \omega, \alpha \sim TN_{(-\infty,\delta)} \left(\mu + \psi \frac{\exp(\xi\alpha_t) - 1}{\xi}, \sigma^2 \right),$$

where $\mathrm{TN}_{(-\infty,\delta)}$ denotes the truncated normal distribution defined over the domain $(-\infty, \delta)$. This additional step is also applicable to other time-dependent GEV models which we shall consider in the following section.

Several works tackle the uncertainty of the threshold δ within a Bayesian inference (e.g., Tancredi et al., 2006). However, because our main focus is to assess the time-dependence in an extreme value process, we put this issue aside, and assume that δ is set to a suitable value in a particular application.

3. The GEV-MA model

3.1. Model specification

In this section, we consider a GEV model with a different kind of dynamic. Instead of the AR(1) process considered in the previous section, time-dependence is incorporated through a state variable following a first order MA process. The resulting model, GEV-MA, for short, combines the measurement equation (5) with the state equation

$$\alpha_{t+1} = \eta_t + \theta\eta_{t-1}, \quad \eta_t \sim G,\ t = 1, \ldots, n - 1, \tag{13}$$

where $|\theta| < 1$. The state variables $\{\alpha_t\}_{t=1}^n$ are assumed to follow an invertible MA(1) process with Gumbel-distributed innovations. The parameter θ measures the degree of dependence in the GEV-MA model. If both θ and σ^2 are zero, the model reduces to i.i.d. observations from the GEV distribution as before. For the initial state α_1, we assume that

$$\alpha_1 = \theta c_0 + \eta_0 + \theta\sqrt{c_1}\eta_0^*, \quad \eta_0^* \sim N(0, 1), \tag{14}$$

where the Gumbel random variable η_{-1} in $\alpha_1 = \eta_0 + \theta\eta_{-1}$ is replaced by a normal random variable with the same mean and variance for simplicity.

3.2. Bayesian estimation

The unknown model parameters of the GEV-MA model are equal to $\omega \equiv (\lambda, \sigma^2, \theta)$. We develop the Bayesian approach based on assuming prior independence between λ, σ^2 and θ, and use the same priors for λ and σ^2 as in Section 2.2. We assume that $(\theta + 1)/2$ follows a Beta prior distribution.

As in Section 2.2, we use the normal mixture distribution defined in (8) to approximate the exact probability density function $g(\eta_t)$ for the state equation (13). Conditional on the mixture indicator variables, $s = \{s_t\}_{t=0}^{n-1}$, the initial distribution (14) and the state equation (13) read for $t = 1, \ldots, n - 1$:

$$\alpha_{t+1} = (m_{s_t} + v_{s_t}u_t) + \theta(m_{s_{t-1}} + v_{s_{t-1}}u_{t-1}), \quad u_t \sim N(0, 1), \tag{15}$$

$$\alpha_1 = \theta c_0 + (m_{s_0} + v_{s_0}u_0) + \theta\sqrt{c_1}\eta_0^*, \tag{16}$$

where $u_0 \sim N(0, 1)$. Conditional on s, Eqs. (5), (15) and (16) form the nonlinear Gaussian state space model.

To perform the MCMC estimation, we regard the mixture indicator variables, $s = \{s_t\}_{t=0}^{n-1}$, and the disturbances, $u = \{u_t\}_{t=0}^{n-1}$, as missing data. To draw a sample from the full posterior distribution, $\pi(\omega, u, s|y)$, we implement the following MCMC algorithm.

Algorithm 2 (*MCMC Algorithm for the GEV-MA Model*)**.**

1. Generate $(\mu, \psi, \xi) \mid \sigma^2, \theta, u, y$.
2. Generate $\sigma^2 \mid \mu, \psi, \xi, \theta, u, y$.
3. Generate $\theta \mid \mu, \psi, \xi, \sigma^2, s, u, y$.
4. Generate $s \mid \omega, u, y$.
5. Generate $u \mid \omega, s, y$.

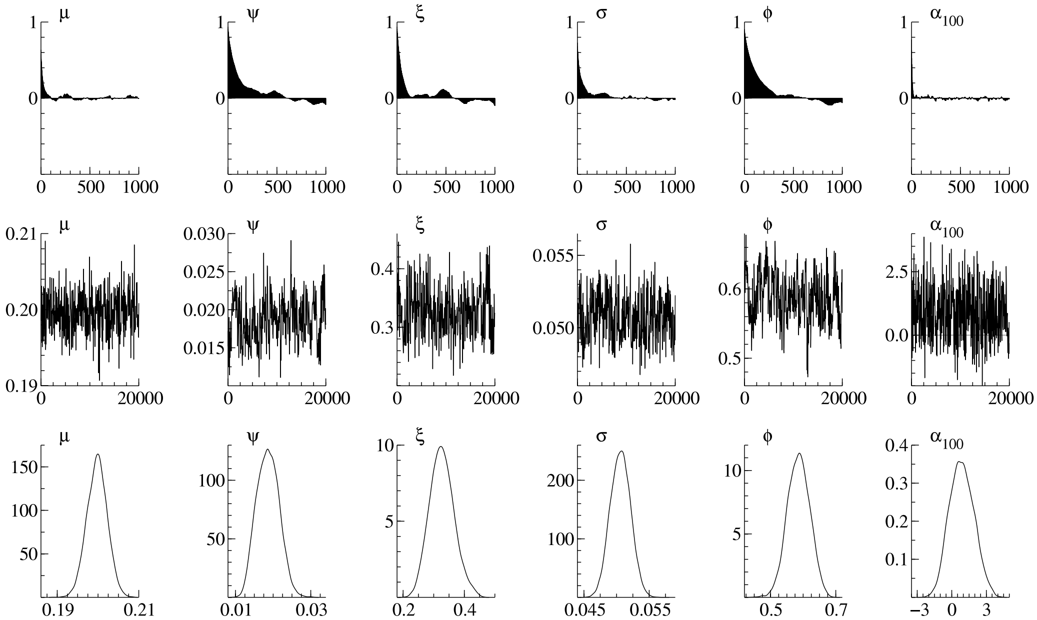

Fig. 1. GEV-AR model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

Steps 1 to 3 are implemented primarily as in Section 2.2. For the generation of θ, however, note that we do not marginalize the conditional posterior distribution over s. Step 4 also differs from the generation for the GEV-AR model because s is no longer conditionally independent of ω, u, and y. This dependence enters through the distribution of α_t in the conditionally Gaussian state model (15), which depends on s_{t-1} and s_{t-2}. Therefore, s_t affects not only the distribution of y_{t+1} as in Section 2.2, but also the distribution of y_{t+2}. The conditional posterior probability mass function of s_t depends on the neighboring values s_{t-1} and s_{t+1}. To make the generation of s_t more efficient, the posterior probability mass function of s_t is marginalized over s_{t+1}, *i.e.*, we sample from the conditional posterior density, $\pi(s_t|\omega, u_{t-1}, u_t, u_{t+1}, u_{t+2}, s_{t-1}, s_{t+2}, y_{t+1}, y_{t+2})$. In Step 5, the disturbance u is generated to obtain the state variable α. We again apply the block sampler developed by Omori and Watanabe (2008) for nonlinear Gaussian state space models. Details for all sampling step are provided in Appendix A.2.

4. Illustrative simulation study

In this section we illustrate the proposed algorithm using simulated data. We generate 2,000 observations from the GEV-AR and the GEV-MA model, respectively, with $\mu = 0.2, \psi = 0.02, \xi = 0.3, \sigma = 0.05, \phi = 0.6$, and $\theta = 0.3$. The following prior distributions are assumed: $\mu \sim N(0, 10), \psi \sim \text{Gamma}(2, 2), \xi \sim N(0, 4), \sigma^2 \sim \text{IG}(2.5, 0.025)$, $(\phi + 1)/2 \sim \text{Beta}(4, 4)$, and $(\theta + 1)/2 \sim \text{Beta}(4, 4)$. We draw $M = 20{,}000$ samples after the initial 10,000 samples are discarded as the burn-in period.

Table 2 reports the estimated posterior means, standard deviations, 95% credible intervals and inefficiency factors. The inefficiency factor is defined as $1+2\sum_{s=1}^{\infty}\rho_s$ where ρ_s is the sample autocorrelation at lag s. It measures how well the MCMC chain mixes (see e.g., Chib, 2001). It is the ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. The inverse of inefficiency factor is also known as relative numerical efficiency (Geweke, 1992). When the inefficiency factor is equal to m, we need to draw MCMC sample m times as many as uncorrelated sample. In the following analyses, we compute the inefficiency factor using a bandwidth $b_w = 1{,}000$.

The estimation result shows that all estimated posterior means are close to the true values and the true values are contained in the 95% credible intervals. Interestingly, the inefficiency factors of the GEV-MA model are considerably lower than for the GEV-AR model. We also report the result of sampling the state variable, α_{100}. The inefficiency factor for α_{100} is relatively low for both models, which indicates that efficient sampling for the state variables is enhanced by the multi-move block sampler.

Figs. 1 and 2 show the sample autocorrelation functions, the sample paths and the posterior densities for the GEV-AR and the GEV-MA model, respectively. The sample paths look stable and the sample autocorrelations drop smoothly.

As described in Section 2.2.2, the difference between draws from the exact and the approximate posterior density can be evaluated through the weight w_j defined in (11). If the approximation is good, we expect the log-weights $\{\log(w_j \times M)\}_{j=1}^{M}$ to follow a distribution with mean 0 and small variance (Kim et al., 1998). Fig. 3 plots the histogram of the $\log(w_j \times M)$ of the GEV-AR and the GEV-MA models with normal density functions setting the mean and variance equal to the individual sample mean and sample variance. The log-weights are concentrated around zero with small variance, which indicates a good approximation of the mixture of normals.

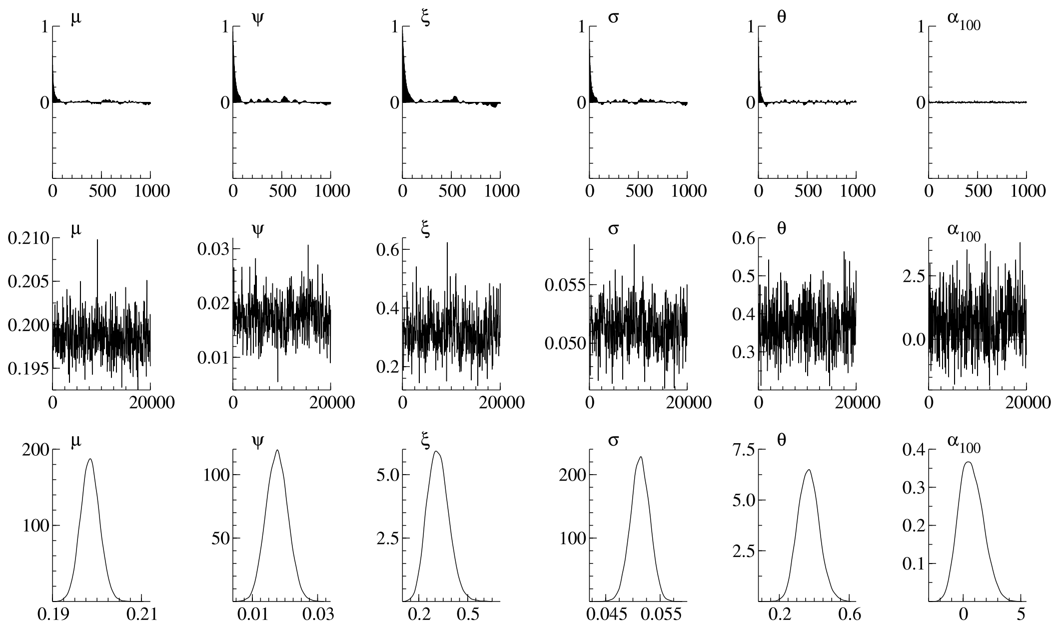

Fig. 2. GEV-MA model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

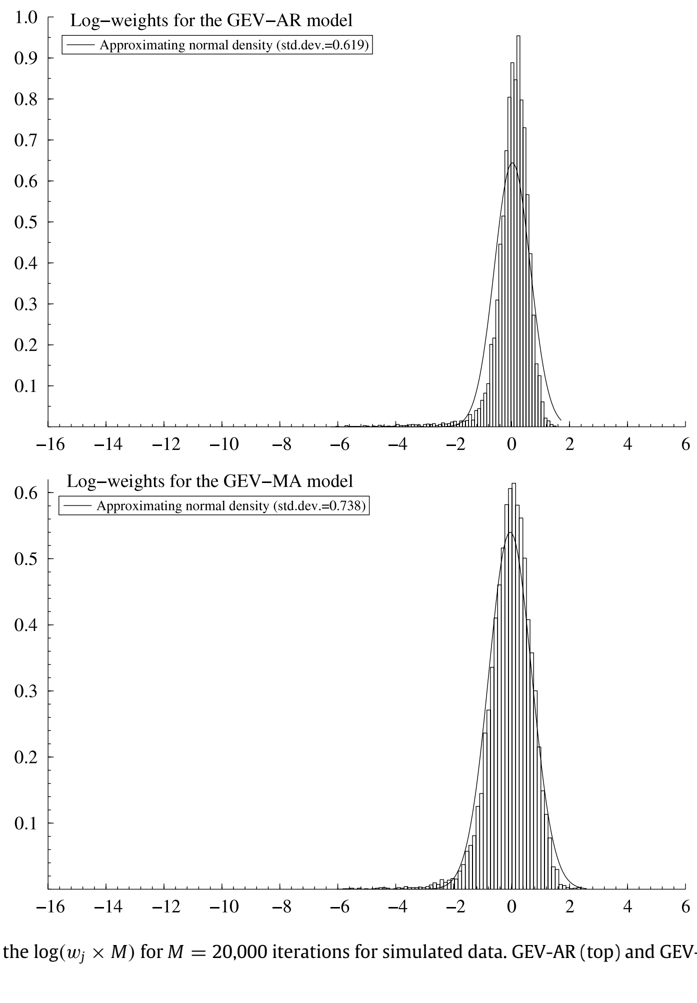

Fig. 3. Histogram of the $\log(w_j \times M)$ for $M = 20{,}000$ iterations for simulated data. GEV-AR (top) and GEV-MA (bottom) models.

Table 2
Estimation result of the GEV models for simulated data. Ineff refers to the inefficiency factors.

Parameter	True	Mean	Stdev.	95% interval	Ineff.
(i) GEV-AR model					
μ	0.2	0.1994	0.0025	[0.1942, 0.2041]	33.5
ψ	0.02	0.0184	0.0030	[0.0132, 0.0244]	253.8
ξ	0.3	0.3247	0.0425	[0.2433, 0.4150]	120.3
σ	0.05	0.0506	0.0015	[0.0476, 0.0534]	99.3
ϕ	0.6	0.5908	0.0336	[0.5283, 0.6543]	270.6
α_{100}	0.15	0.8909	1.0344	[−1.0807, 2.9204]	20.2
(ii) GEV-MA model					
μ	0.2	0.1986	0.0021	[0.1944, 0.2028]	16.7
ψ	0.02	0.0175	0.0034	[0.0111, 0.0246]	34.8
ξ	0.3	0.3186	0.0685	[0.1948, 0.4671]	39.6
σ	0.05	0.0514	0.0018	[0.0477, 0.0547]	33.3
θ	0.3	0.3672	0.0611	[0.2523, 0.4895]	16.0
α_{100}	1.29	0.7029	1.0276	[−1.1152, 2.8223]	1.1

Table 3
Summary statistics for the TOPIX minima data (multiplied by -1, $n = 216$).

Mean	Stdev.	Skewness	Kurtosis	Max.	Min.
2.289	1.201	1.266	4.703	7.100	0.554

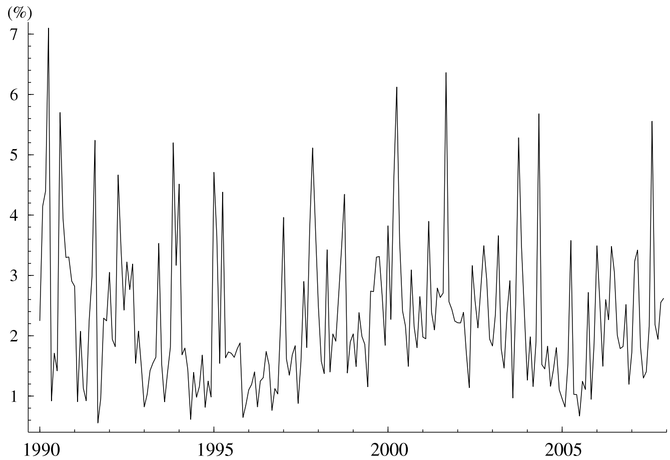

Fig. 4. Minimum return data for the TOPIX (multiplied by −1, 1990/Jan–2007/Dec).

As a side remark, when ξ is close to zero, there could be some confounding between ξ and σ in generating sample, because $\exp(\xi\alpha_t) \approx 1 + \xi\alpha_t + \xi^2/2\alpha_t^2$, and $y_t \approx \mu + \psi\alpha_t + \psi\xi/2\alpha_t^2 + \varepsilon_t$. We checked this point using simulated data by setting $\xi = 10^{-2}$ and 10^{-3}. In result, the scatter plot of ξ and σ exhibited no problematic strong correlations.

5. Application to stock returns data

5.1. Data

In this section, we apply our models to minimum daily stock returns occurring during a month using the Tokyo Stock Price Index (TOPIX). The original sample period is from January 4, 1990 to December 28, 2007. We take log-differences (multiplied by 100) to compute the daily return and pick up the monthly minima, which leads to a series of 216 observations. For estimation, we use the minima multiplied by −1. Table 3 summarizes the descriptive statistics and Fig. 4 shows the time series plot. The skewness is positive and the kurtosis is larger than that of a normal distribution, which implies a longer right tail and fatter tails.

While there are several frequencies to analyze financial market variables, the daily stock return is one of the most popular figures that market participants much care about. The extreme values in their left tail are of interest for a wide

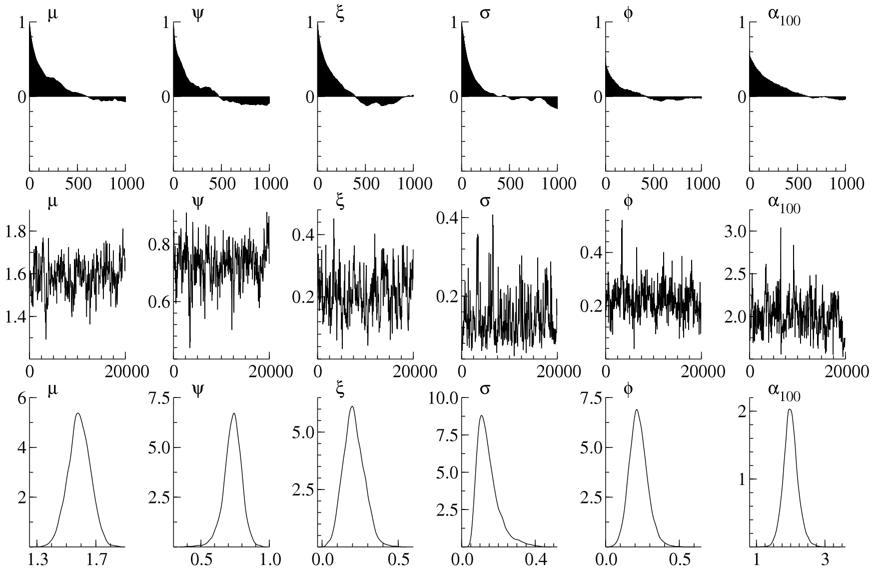

Fig. 5. Estimation result of the GEV-AR model for the TOPIX minima data.

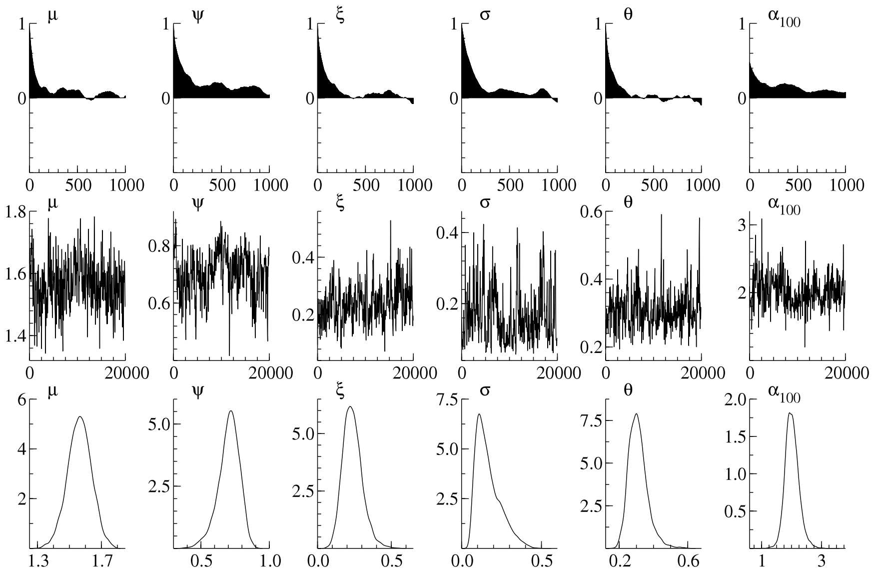

Fig. 6. Estimation result of the GEV-MA model for the TOPIX minima data.

range of applications such as risk management and portfolio selection. In practice, to increase the knowledge about the monthly minima distribution and underlying behavior of the state variables may contribute to the risk control of a large-scale portfolio in financial institutions as well as to the market stability maintenance of monetary authorities. For instance, it is probably plausible to assume that companies and financial institutions may review their asset allocation in monthly basis toward a possible maximum loss of a daily transaction in the next month. From these perspectives, the GEV-AR and GEV-MA models are fit to the monthly minima return data in the following analysis.

5.2. Estimation results

We first estimate three models for the TOPIX minima data; the GEV-AR model, the GEV-MA model and the simple GEV model where $\phi = \theta = 0$, labeled GEV, for short. The prior specifications and the iteration sizes are the same as in the simulation study in Section 4. Figs. 5 and 6 show the sample autocorrelations, sample paths and posterior densities of the GEV-AR and the GEV-MA model respectively for the TOPIX minima data. The MCMC results show that the Markov chains mix well.

Table 4
Estimation result of the GEV models for the TOPIX minima data.

Parameter	GEV	GEV-AR	GEV-MA
μ	1.6987 (0.0680) [1.5709, 1.8371] 216.8	1.5817 (0.0763) [1.4312, 1.7329] 190.0	1.5495 (0.0789) [1.3949, 1.7018] 141.6
ψ	0.8042 (0.0546) [0.7005, 0.9149] 226.1	0.7308 (0.0728) [0.5668, 0.8609] 186.7	0.6912 (0.0746) [0.5197, 0.8249] 181.6
ξ	0.1519 (0.0714) [0.0328, 0.3237] 422.8	0.2115 (0.0647) [0.0938, 0.3470] 185.3	0.2212 (0.0629) [0.1100, 0.3557] 171.2
σ	0.1052 (0.0324) [0.0605, 0.1849] 66.7	0.1443 (0.0648) [0.0657, 0.3334] 197.3	0.1714 (0.0754) [0.0678, 0.3436] 185.0
ϕ		0.2263 (0.0629) [0.1152, 0.3652] 87.3	
θ			0.3080 (0.0548) [0.2168, 0.4327] 144.5
α_{100}	1.8575 (0.1944) [1.4929, 2.2971] 338.3	2.0018 (0.2186) [1.1659, 2.4905] 161.3	2.0851 (0.2433) [1.6242, 2.5965] 116.3

The first row: posterior mean and standard deviation in parentheses.
The second row: 95% credible interval in square brackets.
The third row: inefficiency factor.

Table 4 reports the estimation result of the three GEV models. Regarding the posterior means for the parameters in the GEV distribution, μ and ψ become smaller while ξ turns to be larger in the order of the GEV, GEV-AR and GEV-MA models. Also, the posterior mean of σ becomes higher in this order, which implies that the idiosyncratic error tends to be larger. Concerning the parameters capturing time-dependence, the posterior mean of ϕ is about 0.2 for the GEV-AR model and the posterior mean of θ is about 0.3 for the GEV-MA model. For both models, the 95% credible intervals of the corresponding parameter do not contain zero. These results suggest evidence of time-dependence in the minimum returns, and of both autoregressive and moving average effects in the proposed models.

Regarding the estimates of the parameter ξ, the posterior means are estimated to be positive and the 95% credible intervals do not contain zero for all models. When we consider random variables following a certain distribution function F and the limit distribution of the rescaled maximum is H_ξ, we say that the distribution F lies in the maximum domain of attraction of H_ξ. For $\xi > 0$, the GEV distribution forms the Fréchet distribution and its domain of attraction includes distributions such as the Student-t, the Pareto and the inverse gamma distributions. These distributions have heavier tails, so-called power tails. The estimates of the parameter ξ obtained above imply that the underlying daily return data would follow such a heavy-tailed distribution, as often pointed out in the financial literature.

Fig. 7 plots the histogram of the $\log(w_j \times M)$ of the GEV-AR and GEV-MA models for the TOPIX minima data. The log-weights are concentrated around zero with a small variance, indicating that our approximation based on the mixture of normals is evidently accurate.

5.3. Posterior predictive analysis

To check the plausibility of our proposed model, we conduct a posterior predictive analysis (see, e.g., Gelman et al., 2003, Chapter 6). We generate a set of $n = 216$ new observations for each MCMC draw and calculate for each dataset twelve summary statistics: the sample mean and the sample standard deviation, the median, the lower and the upper quartile, the minimum and the maximum, the sample autocorrelation function (ACF) for the lags 1–3, and the sample partial autocorrelation function (PACF) for the lags 2 and 3. Using the posterior predictive distributions of these statistics, we can check whether the statistics calculated for the originally observed data were likely to occur under the proposed models. The failure to replicate the observed statistics suggests the implausibility of the model.

Fig. 8 shows the density plots of these summary statistics for three competing models where the vertical lines correspond to the actual quantities calculated from the originally observed data. The p in parenthesis denotes the posterior predictive p-value which is the area to the right of the actual statistics. All p-values except the first order ACF (ACF-1) for the GEV and the GEV-MA models assure the plausibility of the model. The small p-values (less than 0.05) for the ACF-1's for these two models imply that they are not entirely replicating the time-dependence present in our data which exhibit the substantial first order autocorrelation.

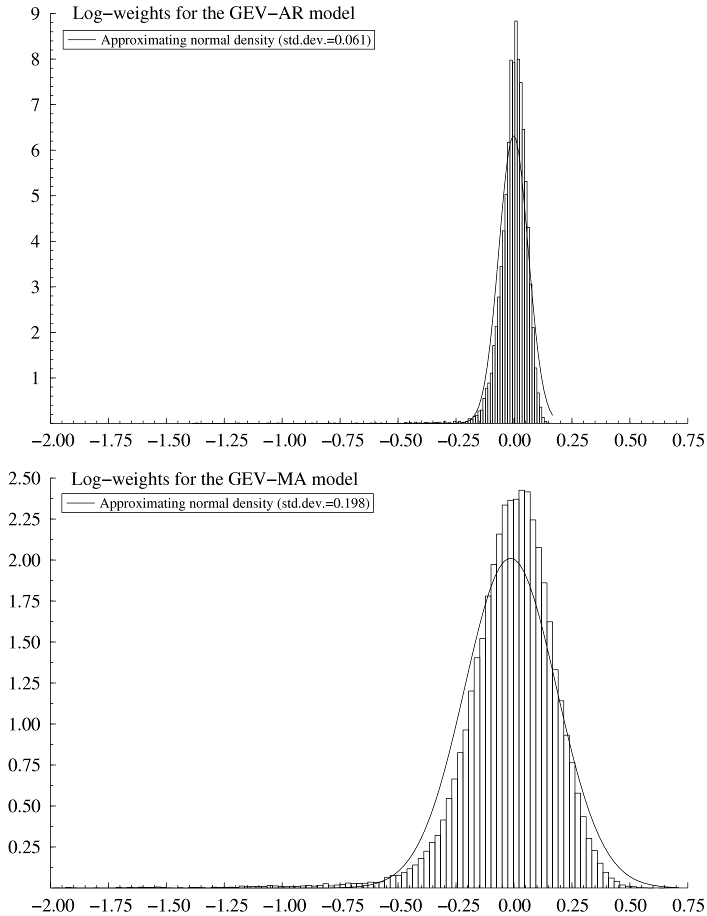

Fig. 7. Histogram of the $\log(w_j \times M)$ for $M = 20{,}000$ iterations for TOPIX minima data. GEV-AR (top) and GEV-MA (bottom) models.

5.4. Model comparisons

This section provides the model comparisons of the proposed model to standard time series models and other dynamic GEV models. The comparison is based on the marginal likelihoods and forecasting performance using the TOPIX minima data.

5.4.1. Competing models

In addition to the GEV, GEV-AR and GEV-MA models, we examine the following models discussed in existing literature. First, as standard time series models, an AR(1) model: $y_t = c + \phi y_{t-1} + \varepsilon_t$; and a MA(1) model: $y_t = c + \varepsilon_t + \theta\varepsilon_{t-1}$ are considered, where the innovations follow a normal distribution, namely, $\varepsilon_t \sim N(0, \sigma^2)$. We assume that $|\phi| < 1$ and $|\theta| < 1$. Second, following Hughes et al. (2007), the AR(1) and MA(1) models with the GEV innovations are investigated, where ε_t is an i.i.d. random variable from the GEV distribution defined in (1), and $c = 0$ in the above equations. We label these models as AR-GEV and MA-GEV, respectively.

Third, following Gaetan and Grigoletto (2004) and Huerta and Sansó (2007), the conventional time-varying parameter GEV models are examined. Specifically, we assume that $y_t \sim \mathrm{GEV}(\mu_t, \psi_t, \xi_t)$, and allow each time-varying parameter of $(\mu_t, \lambda_t(\equiv \log(\psi_t)), \xi_t)$ to follow the AR(1) or MA(1) process. The conventional GEV-AR (labeled as CGEV-AR) model is defined with the state process:

$$\varphi_{t+1} = \mu_\varphi + \phi_\varphi \varphi_t + \varepsilon_{\varphi,t}, \quad \varepsilon_{\varphi,t} \sim N(0, \sigma_\varphi^2),$$

where $\varphi_t = \mu_t, \lambda_t$, and ξ_t. As a similar model, we consider the modified CGEV-AR model where only μ_t is time-varying, *i.e.*, $\psi_t \equiv \psi$ and $\xi_t \equiv \xi$, labeled as CGEV-ARm model. By assuming that $\phi_\varphi = 1$, the CGEV models with random-walk process (CGEV-RW and CGEV-RWm) are also considered. The conventional GEV-MA (labeled as CGEV-MA) model is defined with the state process:

$$\varphi_{t+1} = \mu_\varphi + \varepsilon_{\varphi,t} + \theta_\varphi \varepsilon_{\varphi,t-1}, \quad \varepsilon_{\varphi,t} \sim N(0, \sigma_\varphi^2).$$

We refer the CGEV-MA model where only μ_t is time-varying as CGEV-MAm model.

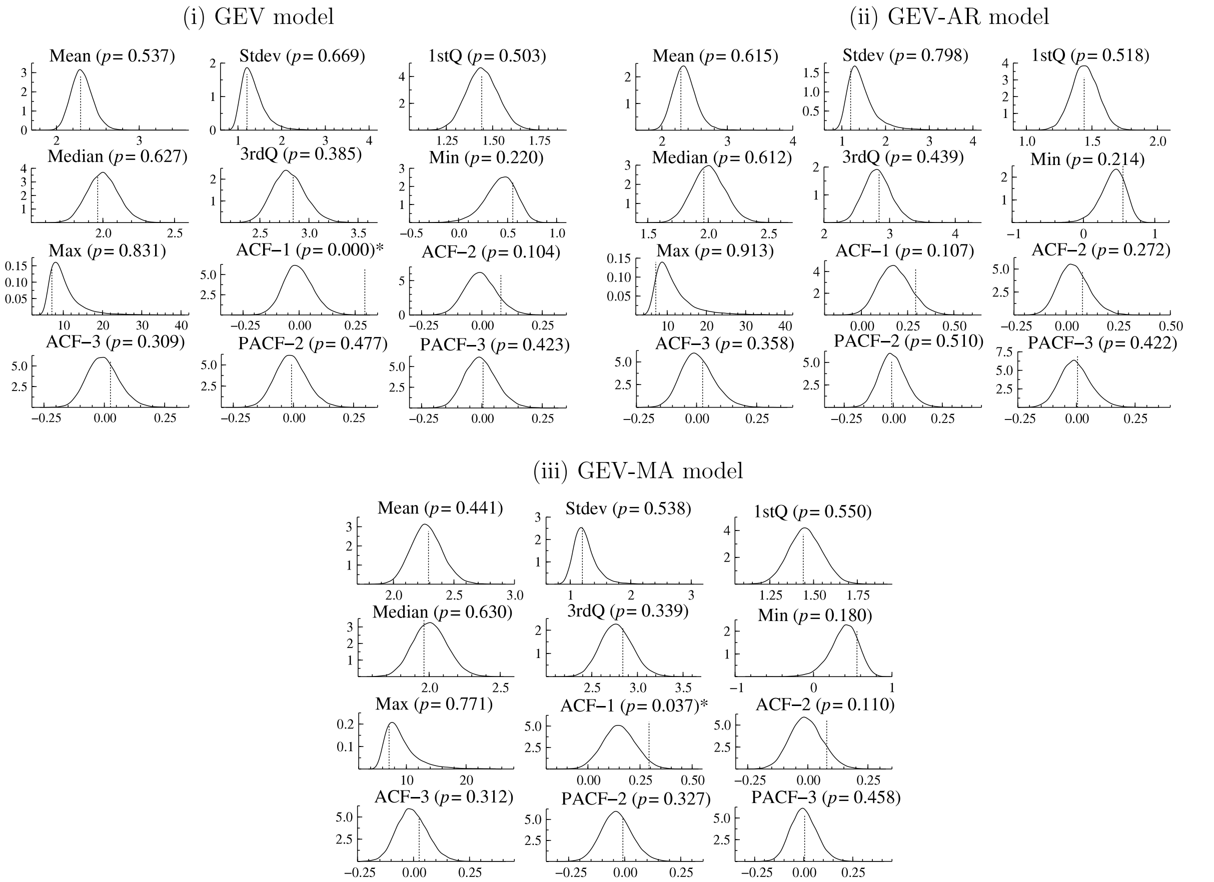

Fig. 8. Posterior predictive check. The vertical lines refers to the values of the test quantities based on the observed data, and the posterior predictive p-value are in parentheses. ($*$) indicates a statistical significance at the 10% significant level.

Last, the threshold models introduced in Section 2.4 are compared for GEV, GEV-AR, GEV-MA models. For estimation we set the threshold δ equal to 1.48 (the 90th percentile of all daily returns multiplied by -1), and use only the minima that exceed the threshold.

5.4.2. Result of marginal likelihoods

The competing models are compared based on the marginal likelihood $m(y)$, which is defined as the integral of the likelihood function with respect to the prior density of the parameters. When the prior probabilities of the competing models are assumed to be equal, we choose the model which yields the largest marginal likelihood. Following Chib (1995), we estimate the log of marginal likelihood using the identity

$$\log m(y) = \log f(y|\Theta) + \log \pi(\Theta) - \log \pi(\Theta|y),$$

where Θ is a parameter set in the model, $f(y|\Theta)$ is the likelihood function, $\pi(\Theta)$ is the prior probability density and $\pi(\Theta|y)$ is the posterior density. This equality holds for any Θ, but we usually use the posterior mean of Θ to obtain a stable estimate of $m(y)$. To evaluate the posterior ordinate $\pi(\Theta|y)$, we use the method of Chib (1995) and Chib and Jeliazkov (2001) using 10,000 draws obtained through reduced MCMC runs. For the GEV models the likelihood function is computed by the particle filters developed in Section 2.3 using $I = 10{,}000$ particles.

The marginal likelihood may be estimated every MCMC iteration based on the draws of Θ, although we have considerable computation burden in the computation of particle filter for the GEV models. Therefore, we calculate the marginal likelihood once after the MCMC iterations based on the posterior mean of Θ. We consider that the marginal likelihood estimate does not substantially differ between these two cases. Ten replications of the filter are implemented to obtain the standard error of the likelihood. We compute the standard error of the posterior ordinate using the method of Chib and Jeliazkov (2001), which does not require the computation of the marginal likelihood every MCMC iteration.

Table 5 report the results of marginal likelihood estimation. The proposed GEV-AR model yields the highest marginal likelihood. The estimated marginal likelihoods of the standard AR(1) and MA(1) models are smaller than that of the other models. This implies the considerable improvement by modeling extreme values using the GEV distributions. Among the conventional GEV models, the CGEV-ARm model is favored over the other models. Interestingly, the models where only

Table 5
Estimated marginal likelihoods (ML) of the competing models for the TOPIX minima data.

	GEV	GEV-AR	GEV-MA	AR(1)	MA(1)	AR-GEV	MA-GEV
Log-ML	−323.75	−318.11	−318.68	−354.81	−355.23	−322.70	−322.91
(S.E.)	(0.39)	(0.57)	(0.58)	(0.001)	(0.001)	(0.01)	(0.01)
	CGEV-AR	-ARm	-RW	-RWm	-MA	-MAm	
Log-ML	−324.60	−320.67	−324.22	−321.75	−323.84	−322.45	
(S.E.)	(0.28)	(0.17)	(0.45)	(0.11)	(0.04)	(0.03)	

	Threshold models		
	GEV	GEV-AR	GEV-MA
Log-ML	−329.05	−318.47	−321.58
(S.E.)	(0.64)	(0.94)	(0.85)

∗ All values are in natural log scale. Standard errors are in parentheses.

Table 6
Mean absolute percentage errors (MAPE) and root mean squared percentage errors (RMSPE) for the posterior median of predictive distribution.

Model	GEV	GEV-AR	GEV-MA	AR(1)	MA(1)	AR-GEV	MA-GEV
MAPE	0.342	0.249	0.261	0.298	0.282	0.282	0.271
RMSPE	0.391	0.310	0.316	0.367	0.359	0.337	0.324
Model	CGEV-AR	-ARm	-RW	-RWm	-MA	-MAm	
MAPE	0.353	0.377	0.322	0.315	0.272	0.318	
RMSPE	0.506	0.457	0.391	0.379	0.327	0.402	

μ_t is time-varying outperform the models where all three parameters are time-varying. Overall, our proposed GEV-AR and GEV-MA models outperform the other dynamic GEV models discussed in existing literature. Table 5 also reports the marginal likelihoods for the threshold models of GEV, GEV-AR and GEV-MA models. The time-dependent models still outperform the GEV model. From these results, the GEV-AR model is found to be the best model for the TOPIX returns minima data among the competing models based on the marginal likelihoods.

5.4.3. Result of forecasting performance

We investigate the second model comparison based on forecasting performance. Consider the one-step ahead predictive density given data y, $\pi(y_{n+1}|y)$,

$$\pi(y_{n+1}|y) = \iiint f(y_{n+1}|y, \omega, s, \alpha)\pi(\omega, s, \alpha|y)d\omega\, ds\, d\alpha.$$

A random sample from this predictive distribution is obtained in the MCMC algorithm by adding one more step to generate $y_{n+1}^i \sim f(y_{n+1}|y, \omega^i, s^i, \alpha^i)$ for the ith iteration given the current sample of parameters and latent variables $(\omega^i, s^i, \alpha^i)$.

To compare the time-dependent extreme value models and other time series models, we compute the posterior median of the one-step ahead predictive distribution for 24 subsample datasets of the stock return minima data. Each dataset includes 192 monthly observations (*i.e.*, 16 years). We first use the sample period from January 1990 to December 2005 to estimate parameters using MCMC, and generate samples from the one-step ahead predictive distribution to forecast the minimum daily stock return in January 2006. Next we use the sample period from February 1990 to January 2006 to predict the minimum return in February 2006. We repeat this rolling estimation until we obtain 24 forecasted values (*i.e.*, 2 years). The number of the MCMC iterations and the prior settings are the same as those in the preceding estimations.

The comparison is based on the mean absolute percentage errors (MAPE) and root mean squared percentage errors (RMSPE), namely,

$$\text{MAPE} = \frac{1}{\tau}\sum_{i=1}^{\tau}\left|\frac{\hat{y}_{N+i} - y_{N+i}}{y_{N+i}}\right|, \qquad \text{RMSPE} = \left\{\frac{1}{\tau}\sum_{i=1}^{\tau}\left(\frac{\hat{y}_{N+i} - y_{N+i}}{y_{N+i}}\right)^2\right\}^{1/2},$$

where $N = 192$, $\tau = 24$, and y_{N+i}, $\hat{y}_{N+i}$ denote the actual value and the posterior predictive median at period $N + i$ using the subsample data from period i to $N + i - 1$.

Table 6 shows the MAPE and RMPSE for the competing GEV and other time series models. The proposed GEV-AR and GEV-MA models clearly outperforms the other competing models based on both MAPE and RMSPE. Because we found evidence of time-dependence in the posterior estimates in Section 5.2, the assumption of the time-dependence evidently contribute to the forecasting performance. The AR-GEV and MA-GEV models perform slightly better than the normal-error models, the AR(1) and MA(1). This implies that the GEV-distributed innovations are favored in forecasting. Among the dynamic CGEV models, the CGEV-RWm, CGEV-MA, and CGEV-MAm models yield relatively better forecasts than the others. However, the CGEV models are dominated by the GEV-AR and GEV-MA models. Although a forecasting performance may

Table 7
Estimated log-likelihoods for the TOPIX minima data using four particle filter methods: New (proposed filter), PF (simple particle filter), APF (Auxiliary particle filter) and MPF (the modified APF).

	GEV-AR				GEV-MA			
	New	PF	MPF	APF	New	PF	MPF	APF
$R = 10{,}000$	−304.39 (0.10)	−304.41 (0.13)	−310.66 (0.84)	−341.49 (1.18)	−306.21 (0.09)	−306.63 (0.10)	−321.81 (0.89)	−367.26 (1.80)
$R = 50{,}000$	−304.28 (0.03)	−304.36 (0.05)	−306.85 (0.52)	−330.23 (1.58)	−306.30 (0.03)	−306.33 (0.08)	−316.61 (0.89)	−351.83 (1.52)
$R = 100{,}000$	−304.33 (0.03)	−304.50 (0.09)	−305.67 (0.61)	−327.34 (1.23)	−306.27 (0.02)	−306.41 (0.06)	−312.55 (0.71)	−348.08 (1.03)

	GEV		
	New	PF	MPF
$R = 10{,}000$	−312.23 (0.08)	−312.20 (0.26)	−312.47 (0.16)
$R = 50{,}000$	−312.01 (0.04)	−312.14 (0.06)	−312.29 (0.07)
$R = 100{,}000$	−312.10 (0.02)	−312.12 (0.06)	−312.17 (0.03)

∗ All values are in natural log scale, and standard errors are in parentheses. R denotes the number of particles.

depend on the evaluated periods of sample, the proposed GEV-AR model performs the best for the prediction of the stock return minima data.

5.5. Comparison of particle filters

We examine the particle filters discussed in Section 2.3 using the number of particles $I = 10{,}000$, 50,000 and 100,000 with the number of replications $M = 10$. For comparison, we also consider the modified APF (MPF), which is similar to the filter proposed by Chamú Morales (2005) in the context of moving maxima processes. The MPF is the filter for the observation y_{t+1} less than a certain threshold, say, the 95th percentile of y_t's. For the extreme observation y_{t+1} which exceeds the threshold, it employs the importance function based on a mixture distribution, $q(\alpha_{t+1}|y_{t+1}, \alpha_t, \omega) = 0.95g(\alpha_{t+1}|\alpha_{t+1} > u_{t+1}, \omega) + 0.05f(\alpha_{t+1}|\alpha_t, \omega)$, where $g(\alpha_{t+1}|\alpha_{t+1} > u_{t+1}, \omega)$ is the truncated Gumbel density with the location $\phi\alpha_t$. The truncation point u_{t+1} is chosen so that m_{t+1} is equal to a median of the truncated Gumbel distribution.

Table 7 reports the estimated likelihoods for the TOPIX minima data. The APF is compared only in the GEV-AR and GEV-MA models. For the GEV-AR and GEV-MA models, it is clear that our new filter produces more stable and accurate estimates than the APF and the MPF. The APF tends to yield the unstable estimates in the existence of extreme observations. The standard PF produces stable estimates, but their standard errors are larger than those of the new filter. For the GEV model, the estimates are stable for all the filters, but the standard errors of our new filter are found to be smaller than those of MPF and PF. These estimation results suggest that our proposed method tends to outperform the other filters.

6. Conclusion

In the context of extreme value modeling, this paper develops the new approach to model time-dependence in the GEV distribution using the state space representation where the state variables either follows the AR or an MA processes with innovations from the Gumbel distribution. Approximating the Gumbel density by the ten-component mixture of normal distributions, the mixture sampler is proposed to implement the Markov chain Monte Carlo methods. In the application to the monthly series of minimum daily returns for the TOPIX data, the posterior estimates imply evidence of the underlying time-varying structure in the GEV models. Based on the estimated marginal likelihoods and forecasting performance, the proposed GEV-AR model outperforms other existing time series model and the conventional dynamic GEV models.

In the present paper, we assumed for time-dependent GEV models that the idiosyncratic error ε_t appearing in the observation equation (5) follows a normal distribution for simplicity. One might expect that the distribution of the idiosyncratic error is non-normal, skewed or fat-tailed. The proposed estimation methodology can be extended easily to deal with non-normal errors in the observation equation. For instance, the extension to a skew-t error distribution is straightforward by adding several steps to the developed sampling algorithm. Meanwhile, the posterior predictive analysis provided in Section 5.3 indicated that the assumption of the normal errors tends to be plausible in the sense that it replicates the characteristics of the observations.

Acknowledgments

The authors would like to thank Erricos Kontoghiorghes, the Associate Editor, anonymous referees, Gianni Amisano, Sien Jan Koopman, Hideo Kozumi, Roberto Leon-Gonzalez, W.K. Li, Haruhisa Nishino, Kosuke Oya, Dale J. Poirier, Wolfgang

Polasek, Esther Ruiz, Jun Yu, Philip Yu and seminar participants at the International Conference "High-frequency Data Analysis in Financial Markets" at Hitotsubashi University, and the conference "Recent Developments in Finance and Econometrics" at University of Ryukyus for their helpful comments. This work is supported by Nakajima Foundation, and the Grants-in-Aid for Scientific Research (A) 21243018 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. The computational results are generated using Ox version 4.02 (Doornik, 2006).

Appendix. Details on MCMC estimation

A.1. GEV-AR model

A.1.1. Generation of the model parameters (μ, ψ, ξ), σ^2 and ϕ

In Step 1 of Algorithm 1, the conditional posterior probability density of $\lambda = (\mu, \psi, \xi)'$ is given by $\pi(\lambda|\sigma^2, \alpha, y) \propto \pi(\lambda)f(y|\lambda, \sigma^2, \alpha)$, where f is the conditional likelihood of the observation equation (5). To sample from the conditional posterior distribution, we implement a Metropolis–Hastings (MH) algorithm with following normal proposal density. First we find $\hat{\lambda} = (\hat{\mu}, \hat{\psi}, \hat{\xi})'$ which maximizes (or approximately maximizes) the conditional posterior density. Next we generate a candidate λ^* from a normal distribution truncated over the region $R = \{\psi : \psi \le 0\}$, $\lambda^* \sim \mathrm{TN}_R(\lambda_*, \Sigma_*)$, where

$$\lambda_* = \hat{\lambda} + \Sigma_* \left.\frac{\partial \log \pi(\lambda|\sigma^2, \alpha, y)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}}, \qquad \Sigma_*^{-1} = -\left.\frac{\partial \log \pi(\lambda|\sigma^2, \alpha, y)}{\partial \lambda \partial \lambda'}\right|_{\lambda=\hat{\lambda}},$$

and accept it with probability

$$\alpha(\lambda, \lambda^*) = \min\left\{\frac{\pi(\lambda^*|\sigma^2, \alpha, y)f_N(\lambda|\lambda_*, \Sigma_*)}{\pi(\lambda|\sigma^2, \alpha, y)f_N(\lambda^*|\lambda_*, \Sigma_*)}, 1\right\},$$

where λ denotes the current value and $f_N(\cdot|\mu, \Sigma)$ denotes the probability density function of the normal distribution with mean μ and covariance matrix Σ. If the candidate λ^* is rejected, we take λ as a new sample.

In Step 2, we sample from $\sigma^2|\lambda, \alpha, y \sim \mathrm{IG}(\hat{n}/2, \hat{S}/2)$, where $\hat{n} = n_0 + n$, $\hat{S} = S_0 + \sum_{t=1}^{n}[y_t - \mu - \psi\{\exp(\xi\alpha_t) - 1\}/\xi]^2$.

In Step 3, the conditional posterior density of ϕ is given by

$$\pi(\phi|\alpha) \propto \pi(\phi) \times \pi(\alpha_1|\phi) \times \prod_{t=1}^{n-1}\sum_{j=1}^{K}\pi(\alpha_{t+1}, s_t = j|\phi, \alpha_t).$$

We generate a sample ϕ using the MH algorithm as in Step 1, where the proposal distribution is a truncated normal distribution over the region $|\phi| < 1$.

A.1.2. Generation of s

In Step 4, we simply draw a sample s_t from its discrete conditional posterior distribution with a probability mass function,

$$\pi(s_t = j|\phi, \alpha) \propto p_j \times \frac{1}{v_j}\exp\left\{-\frac{(\alpha_{t+1} - m_j - \phi\alpha_t)^2}{2v_j^2}\right\},$$

for $j = 1, \ldots, K$, and $t = 1, \ldots, n-1$.

A.1.3. Sampling α

In Step 5, we implement a block sampler which divides the state variables α into several blocks and samples each block given other blocks (Shephard and Pitt, 1997; Watanabe and Omori, 2004). To divide $(\alpha_1, \ldots, \alpha_n)$ into $K+1$ blocks, say, $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})$ for $i = 1, \ldots, K+1$ with $k_0 = 0$ and $k_{K+1} = n$, we use the stochastic knots given by $k_i = \mathrm{int}[n(i + U_i)/(K+2)]$, for $i = 1, \ldots, K$, where U_i is a random sample from a uniform distribution $U[0, 1]$ (Shephard and Pitt, 1997). Selecting $(k_1, \ldots, k_K)$ at random for every MCMC iteration would make sampling α more efficient.

To implement the multi-move sampler, we consider the nonlinear state space model

$$\begin{aligned}
&y_t = \mu + \psi\frac{\exp(\xi\alpha_t) - 1}{\xi} + \sigma e_t, \quad t = 1, \ldots, n, \\
&\alpha_{t+1} = w_t + \phi\alpha_t + H_t u_t, \quad t = 0, \ldots, n-1, \\
&(e_t, u_t)' \sim N(0, I_2), \quad t = 1, \ldots, n, \\
&w_t = \begin{cases} \dfrac{c_0}{1-\phi}, & \text{if } t = 0, \\ m_{s_t}, & \text{if } t \ge 1, \end{cases} \qquad H_t = \begin{cases} \sqrt{\dfrac{c_1}{1-\phi^2}}, & \text{if } t = 0, \\ v_{s_t}, & \text{if } t \ge 1, \end{cases}
\end{aligned}$$

where $\alpha_0 = 0$. Let $\vartheta = \left(\omega, \alpha_{r-1}, \alpha_{r+d+1}, \{s_t\}_{t=r-1}^{r+d}, \{y_t\}_{t=r}^{r+d}\right)$. To sample a block $(\alpha_r, \ldots, \alpha_{r+d})$ from its joint conditional posterior distribution, (note that $r \geq 1, d \geq 2, r+d \leq n$), we draw $(u_{r-1}, \ldots, u_{r+d-1})$ whose conditional posterior probability density is

$$\pi(u_{r-1}, \ldots, u_{r+d-1}|\vartheta) \propto \prod_{t=r}^{r+d} \exp\left[-\frac{1}{2\sigma^2}\left\{y_t - \mu - \psi\frac{\exp(\xi\alpha_t)-1}{\xi}\right\}^2\right] \times \prod_{t=r-1}^{r+d-1} \exp\left(-\frac{u_t^2}{2}\right) \times f(\alpha_{r+d}), \tag{A.1}$$

where

$$f(\alpha_{r+d}) = \begin{cases} \exp\left[-\dfrac{(\alpha_{r+d+1} - m_{s_{r+d}} - \phi\alpha_{r+d})^2}{2v_{s_{r+d}}^2}\right], & \text{if } r+d < n, \\ 1, & \text{if } r+d = n, \end{cases}$$

using an MH algorithm. The posterior sample of $(\alpha_r, \ldots, \alpha_{r+d})$ can be obtained by running the state equation using a sample of $(u_{r-1}, \ldots, u_{r+d-1})$ given α_{r-1}. To conduct MH algorithm, we construct the proposal distribution as follows. For $t = r, \ldots, r+d-1$ and $r+d = n$, we consider a Taylor expansion of the logarithm of the likelihood (excluding the constant term)

$$h(\alpha_t) \equiv -\frac{1}{2\sigma^2}\left\{y_t - \mu - \psi\frac{\exp(\xi\alpha_t)-1}{\xi}\right\}^2, \tag{A.2}$$

around the conditional mode $\hat{\alpha}_t$. Let $h'(\hat{\alpha}_t)$ and $h''(\hat{\alpha}_t)$ denote the first and the second derivative of $h(\alpha_t)$ evaluated at $\alpha_t = \hat{\alpha}_t$, respectively. Then,

$$\begin{aligned} h(\alpha_t) &\approx h(\hat{\alpha}_t) + h'(\hat{\alpha}_t)(\alpha_t - \hat{\alpha}_t) + \frac{1}{2}h''(\hat{\alpha}_t)(\alpha_t - \hat{\alpha}_t)^2 \\ &= \frac{1}{2}h''(\hat{\alpha}_t)\left\{\alpha_t - \left(\hat{\alpha}_t - \frac{h'(\hat{\alpha}_t)}{h''(\hat{\alpha}_t)}\right)\right\}^2 + \text{const.} \\ &= -\frac{(y_t^* - \alpha_t)^2}{2\sigma_t^{*2}} + \text{const.}, \end{aligned} \tag{A.3}$$

where $\sigma_t^{*2} = -\{h''(\hat{\alpha}_t)\}^{-1}$ and $y_t^* = \hat{\alpha}_t + \sigma_t^{*2}h'(\hat{\alpha}_t)$ for $t = r, \ldots, r+d-1$ and $t = r+d = n$ and

$$\sigma_{r+d}^{*2} = \frac{1}{-h''(\hat{\alpha}_{r+d}) + \phi^2/v_{s_{r+d}}^2},$$

$$y_{r+d}^* = \sigma_{r+d}^{*2}\left\{h'(\hat{\alpha}_{r+d}) - h''(\hat{\alpha}_{r+d})\hat{\alpha}_{r+d} + \phi(\alpha_{r+d+1} - m_{s_{r+d}})/v_{s_{r+d}}^2\right\},$$

for $t = r+d < n$. As proposal probability density we use

$$q(u_{r-1}, \ldots, u_{r+d-1}|\vartheta) \propto \prod_{t=r}^{r+d} \exp\left\{-\frac{(y_t^* - \alpha_t)^2}{2\sigma_t^{*2}}\right\} \times \prod_{t=r-1}^{r+d-1} \exp\left(-\frac{u_t^2}{2}\right),$$

which is the posterior density of $(u_{r-1}, \ldots, u_{r+d-1})$ obtained from the state space model

$$\begin{aligned} &y_t^* = \alpha_t + \sigma_t^*\zeta_t, \quad t = r, \ldots, r+d, \\ &\alpha_{t+1} = m_{s_t} + \phi\alpha_t + v_{s_t}u_t, \quad t = r-1, \ldots, r+d-1, \\ &(\zeta_t, u_t)' \sim N(0, I_2), \quad t = r, \ldots, r+d. \end{aligned} \tag{A.4}$$

To generate the candidate $(u_{r-1}, \ldots, u_{r+d-1})$ using $q(u_{r-1}, \ldots, u_{r+d-1}|\vartheta)$, we run Kalman filter and the simulation smoother with the current $(y_r^*, \ldots, y_{r+d}^*)$, $(\sigma_r^{2*}, \ldots, \sigma_{r+d}^{2*})$ in (A.4) (e.g. de Jong and Shephard, 1995; Durbin and Koopman, 2002).

The conditional modes $(\hat{\alpha}_r, \ldots, \hat{\alpha}_{r+d})$ can be found by repeating the following steps for several times until the smoothed state variables converge:

1. Initialize $(\hat{\alpha}_r, \ldots, \hat{\alpha}_{r+d})$.
2. Compute $(y_r^*, \ldots, y_{r+d}^*)$, $(\sigma_r^{2*}, \ldots, \sigma_{r+d}^{2*})$.
3. Run Kalman filter and the disturbance smoother (e.g. Koopman, 1993) using the current points $(y_r^*, \ldots, y_{r+d}^*)$, $(\sigma_r^{2*}, \ldots, \sigma_{r+d}^{2*})$ in (A.4) and obtain $\hat{\alpha}_t^* \equiv E(\alpha_t|\vartheta)$ for $t = r, \ldots, r+d$.
4. Replace $(\hat{\alpha}_r, \ldots, \hat{\alpha}_{r+d})$ by $(\hat{\alpha}_r^*, \ldots, \hat{\alpha}_{r+d}^*)$.
5. Go to 2.

A.2. GEV-MA model

A.2.1. Generation of s

In Step 4, the conditional posterior probability mass function of s_t is given by

$$\begin{aligned}
&\pi(s_t|\omega, u_{t-1}, u_t, u_{t+1}, u_{t+2}, s_{t-1}, s_{t+2}, y_{t+1}, y_{t+2}) \\
&\quad \propto \sum_{j=1}^{K} \pi(s_t, s_{t+1} = j|\omega, u_{t-1}, u_t, u_{t+1}, u_{t+2}, s_{t-1}, s_{t+2}, y_{t+1}, y_{t+2}) \\
&\quad \propto \sum_{j=1}^{K} p_{s_t} \times p_j \times \prod_{k=1}^{3} \exp\left[-\frac{1}{2\sigma^2}\left\{y_{t+k} - \mu - \psi\frac{\exp(\xi\alpha_{t+k}) - 1}{\xi}\right\}^2\right],
\end{aligned}$$

for $t = 1, \ldots, n-3$, where

$$\begin{aligned}
\alpha_{t+1} &= (m_{s_t} + v_{s_t}u_t) + \theta(m_{s_{t-1}} + v_{s_{t-1}}u_{t-1}), \\
\alpha_{t+2} &= (m_j + v_j u_{t+1}) + \theta(m_{s_t} + v_{s_t}u_t), \\
\alpha_{t+3} &= (m_{s_{t+2}} + v_{s_{t+2}}u_{t+2}) + \theta(m_j + v_j u_{t+1}).
\end{aligned}$$

Sampling s_0, s_{n-2} and s_{n-1} from their conditional posterior distribution can be implemented similarly.

A.2.2. Sampling u

In Step 5, we implement a multi-move sampler by dividing the disturbance vector u into several blocks. Because the state α_t depends on only u_{t-1} and u_{t-2} (given s and ω), we sample $(u_{t-1}, \ldots, u_{t+d-1})$ given u_{t-2} and u_{t+d} rather than α_{t-1} and α_{t+d+1}. The state space representation of the GEV-MA model is given by

$$\begin{aligned}
y_t &= \mu + \psi\frac{\exp(\xi z\gamma_t) - 1}{\xi} + \sigma e_t, \quad z = (1, 0),\ t = 1, \ldots, n, \\
\gamma_{t+1} &= w_t + T\gamma_t + H_t u_t, \quad t = 1, \ldots, n-1, \\
\gamma_1 &= w_0 + H_0\begin{pmatrix}\eta_0^* \\ u_0\end{pmatrix}, \qquad \gamma_t = \begin{pmatrix}\alpha_t \\ \beta_t\end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
w_t &= \begin{pmatrix}1 \\ \theta\end{pmatrix} m_{s_t}, \qquad T = \begin{pmatrix}0 & 1 \\ 0 & 0\end{pmatrix}, \qquad H_t = \begin{pmatrix}v_{s_t} \\ \theta v_{s_t}\end{pmatrix}, \quad t = 1, \ldots, n-1, \\
w_0 &= \begin{pmatrix}\theta_0 c_0 + m_{s_0} \\ \theta m_{s_0}\end{pmatrix}, \qquad H_0 = \begin{pmatrix}\theta\sqrt{c_1} & v_{s_0} \\ 0 & \theta v_{s_0}\end{pmatrix}, \\
&(e_t, u_t)' \sim N(0, I_2), \quad t = 1, \ldots, n, \qquad (\eta_0^*, u_0)' \sim N(0, I_2).
\end{aligned}$$

The joint conditional posterior density function of $(u_{r-1}, \ldots, u_{r+d-1})$ is

$$\begin{aligned}
\pi(u_{r-1}, \ldots, u_{r+d-1}|\vartheta) \propto{}& \prod_{t=r}^{r+d} \exp\left[-\frac{1}{2\sigma^2}\left\{y_t - \mu - \psi\frac{\exp(\xi z\gamma_t) - 1}{\xi}\right\}^2\right] \\
&\times \prod_{t=r-1}^{r+d-1} \exp\left(-\frac{u_t^2}{2}\right) \times f(y_{r+d+1}|\omega, \alpha_{r+d+1}),
\end{aligned} \tag{A.5}$$

where

$$f(y_{r+d+1}|\omega, \alpha_{r+d+1}) = \begin{cases} \exp\left[-\frac{1}{2\sigma^2}\left\{y_{r+d+1} - \mu - \psi\frac{\exp(\xi\alpha_{r+d+1}) - 1}{\xi}\right\}^2\right], & \text{if } r + d < n, \\ 1, & \text{if } r + d = n, \end{cases} \tag{A.6}$$

and $\vartheta = \left(\omega, u_{r-2}, u_{r+d}, \{s_t\}_{t=r-1}^{r+d}, \{y_t\}_{t=r}^{r+d+1}\right)$. The α_{r+d+1} in (A.6) is obtained from the state equations. To sample $(u_{r-1}, \ldots, u_{r+d-1})$ from its joint conditional posterior distribution using the MH algorithm, we construct the proposal distribution based on the approximate linear Gaussian state space model as in Appendix A.1.3.

References

Chamú Morales, F., 2005. Estimation of max-stable processes using Monte Carlo methods with applications to financial risk assessment. Ph.D. Dissertation. Department of Statistics, University of North Carolina, Chapel Hill.
Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.
Chib, S., 2001. Markov chain Monte Carlo methods: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3569–3649.
Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association 96, 270–291.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.
Deheuvels, P., 1983. Point processes and multivariate extreme values. Journal of Multivariate Analysis 13, 257–272.
de Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.
Doornik, J., 2006. Ox: Object Oriented Matrix Programming. Timberlake Consultants Press, London.
Durbin, J., Koopman, S.J., 2002. Simple and efficient simulation smoother for state space time series analysis. Biometrika 89, 603–616.
Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer, New York.
Frühwirth-Schnatter, S., Frühwirth, R., 2007. Auxiliary mixture sampling with applications to logistic models. Computational Statistics and Data Analysis 51, 3509–3528.
Frühwirth-Schnatter, S., Frühwirth, R., Held, L., Rue, H., 2009. Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. Statistics and Computing 19, 479–492.
Frühwirth-Schnatter, S., Wagner, H., 2006. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. Biometrika 93, 827–841.
Gaetan, C., Grigoletto, M., 2004. Smoothing sample extremes with dynamic models. Extremes 7, 221–236.
Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference, 2nd ed. Chapman & Hall, CRC, Boca Raton, FL.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, 2nd ed. Chapman & Hall, CRC, London.
Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, New York, pp. 169–188.
Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics. Wiley.
Huerta, G., Sansó, B., 2007. Time-varying models for extreme values. Environmental and Ecological Statistics 14, 285–299.
Hughes, G.L., Rao, S.S., Rao, T.S., 2007. Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. Proceedings of the Royal Society Series A 463, 241–259.
Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.
Koop, G., 2003. Bayesian Econometrics. Wiley, Chichester.
Koopman, S.J., 1993. Disturbance smoother for state space models. Biometrika 80, 117–126.
Kunihama, T., Omori, Y., Zhang, Z., 2011. Bayesian estimation and particle filter for max-stable processes. Journal of Time Series Analysis (in press).
Leadbetter, K.R., Lindgren, G., Rootzén, H., 2004. Extremes and related properties of random sequences and series. New York.
Omori, Y., Chib, S., Shephard, N., Nakajima, J., 2007. Stochastic volatility with leverage: fast likelihood inference. Journal of Econometrics 140, 425–449.
Omori, Y., Watanabe, T., 2008. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. Computational Statistics and Data Analysis 52, 2892–2910.
Pitt, M., Shephard, N., 1999. Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 94, 590–599.
Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.
Smith, R.L., 2003. Statistics of extremes, with applications in environment, insurance and finance. In: Finkenstadt, B., Rootzen, H. (Eds.), Extreme Values in Finance, Telecommunications and the Environment. Chapman and Hall, CRC Press, London, pp. 1–78. (Chapter 1).
Smith, R.L., Miller, J.E., 1986. A non-Gaussian state space model and application to prediction of records. Journals of the Royal Statistical Society 48, 79–88.
Smith, R.L., Weissman, I., 1996. Characterization and estimation of the multivariate extremal index. Manuscript.
Tancredi, A., Anderson, C., O'Hagan, A., 2006. Accounting for threshold uncertainty in extreme value estimation. Extremes 9, 87–106.
Watanabe, T., Omori, Y., 2004. A multi-move sampler for estimating non-Gaussian time series models: comments on Shephard & Pitt (1997). Biometrika 91, 246–248.
Zhang, Z., Smith, R.L., 2004. The behavior of multivariate maxima of moving maxima processes. Journal of Applied Probability 41, 1113–1123.