IMES DISCUSSION PAPER SERIES

Generalized Extreme Value Distribution with Time-Dependence Using the AR and MA Models in State Space Form

Jouchi Nakajima, Tsuyoshi Kunihama, Yasuhiro Omori, and
Sylvia Früwirth-Scnatter

Discussion Paper No. 2009-E-32

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

2-1-1 NIHONBASHI-HONGOKUCHO
CHUO-KU, TOKYO 103-8660
JAPAN

IMES Discussion Paper Series 2009-E-32
November 2009

Generalized Extreme Value Distribution with Time-Dependence Using the AR and MA Models in State Space Form

Jouchi Nakajima*, Tsuyoshi Kunihama, Yasuhiro Omori***, and Sylvia Früwirth-Scnatter******

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 a very 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 for time-dependence among the observed minimum returns.

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

JEL classification: C11, C51, G17

*****Economist, Institute for Monetary and Economic Studies, Bank of Japan. Currently in the Personnel and Corporate Affairs Department (studying at Duke University, E-mail: jouchi.nakajima@stat.duke.edu)

******Graduate student, Graduate School of Economics, University of Tokyo. (E-mail: ee097005@mail.ecc.u-tokyo.ac.jp)

*******Professor, Faculty of Economics, University of Tokyo. (E-mail: omori@e.u-tokyo.ac.jp)

********Professor, Department of Applied Statistics, Johannes Kepler University in Lintz. (E-mail: Sylvia.Fruehwirth-Schnatter@jku.at)

The authors would like to thank Sien Jan Koopman, W. K. Li, Esther Ruiz, Jun Yu, Philip Yu, Hideo Kozumi, Wolfgang Polasek, Roberto Leon-Gonzalez, Kosuke Oya, Haruhisa Nishino, 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, and the staff of the Institute for Monetary and Economic Studies (IMES), the Bank of Japan, for their helpful comments. This work is supported by the Grants-in-Aid for Scientific Research (A) 21243018 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. Views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

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 behaviour 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)). Under such an independence assumption, it is straightforward to compute the likelihood function and to obtain the maximum likelihood estimator of unknown model parameters.

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. The 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 marginal distribution equal to the Fréchet distribution. These processes have been extended to the maxima of moving maxima process (Deheuvels (1983)) and the multivariate maxima of moving maxima process (e.g., Smith and Weissman (1996), Zhang and Smith (2004) and Chamú Morales (2005)). Smith (2003) provides some applications of these classes for the extremes of financial data.

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 state variable in 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 $1+\xi(y_t-\mu)/\psi > 0$ which 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. In the case of $\mu = 0$ and $\psi = 1$, the distribution of a Gumbel random variable, α_t, is given by

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

The mean and variance of α_t is $E(\alpha_t) = c_0$ and $\mathrm{Var}(\alpha_t) = c_1 = \pi^2/6$, respectively, where c_0 is the Euler constant. If Y_t follows the GEV distribution defined by (1) and we further 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 defined by (2). This leads to the following relation between Y_t and α_t,

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

where $\alpha_t \sim G$. In this paper, we consider a state space model for extreme values in which the measurement equation is formed as (4) with an additional idiosyncratic shock as in Chamú Morales (2005). Regarding α_t as a state variable, we model the state equation either in form of an autoregressive (AR) process or a moving average (MA) process with the disturbances following the Gumbel distribution.

To the best of our knowledge, this is the first attempt to discuss such a time-dependence in literature. A conventional approach for the time-varying GEV model is to let the parameters in the GEV distribution, which are regarded as state variables, follow a random walk process with normal 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 state variable with the observation equation satisfying the GEV distribution. The key feature of the model proposed in this paper is to introduce a latent stochastic process where the innovations follow a Gumbel error distribution. A similar model is proposed by Hughes et al. (2007), however, they develop a linear time series model using an ARMA representation with innovations following the extreme value distribution.

In principle, a theoretical limit model of an extreme value process is applicable only to some restricted cases under the presence of temporal dependence. But, as suggested by Coles (2001), in applications "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 new model proposed 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 a time-dependence is estimated to exist, then the time-dependence in our model would provide a better approximation to the underlying process than the basic extreme value model assuming time-independence and would be useful to describe the dynamics of the time series and its prediction. Our way of modeling time-dependence in the extreme value model is not ad hoc, but is based on theoretical derivations to the extent that equation (4) holds. From another point of view, the second term in (4) is a Box-Cox transformation of $\exp(\alpha_t)$ which is often considered for non-Gaussian modelling.

Since the time-dependent GEV model takes the form of a nonlinear, non-Gaussian state space model, it is difficult to implement maximum likelihood estimation of the unknown parameters. It would be possible, but computationally intensive, to apply particle filtering to find the ML estimator. Thus, we pursue a Bayesian approach using Markov chain Monte Carlo (MCMC) method (see, e.g., Chib (2001), Koop (2003), Geweke (2005), and Gamerman and Lopes (2006)) for efficient estimation of the time-dependent GEV model.

To facilitate MCMC estimation, we exploit the very accurate approximation of the Gumbel density by a ten-component mixture of normal densities proposed in Frühwirth-Schnatter and Frühwirth (2007). Introducing for each time t the mixture indicator as auxiliary variable reduces the non-Gaussian non-linear state space model to a conditionally

Gaussian non-linear state space model which allows 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 which they approximate a log χ_1^2 density by mixture of normal densities in the context of stochastic volatility models.

The rest of paper is organized as follows. Section 2 defines a GEV model where the state variables follow an AR(1) process and develops an MCMC algorithm for estimation. In addition, an efficient particle filter is proposed to compute the likelihood function. Furthermore, the model is extended to a threshold model where observations are observed only when they exceed a certain fixed value. Sections 3 introduces the GEV model where the state variables follow an MA(1) process and discusses an appropriate MCMC algorithm. Section 4 illustrates our estimation procedure using simulated data. In Section 5, we apply our method to extreme returns of daily stock data and provide a posterior predictive analysis, model comparisons and forecasting performance comparisons. Section 6 concludes.

2 The GEV-AR model

2.1 Model specification

Let $y = \{y_1, \ldots, y_n\}$ be a sequence of extreme values and G denote the Gumbel distribution given in (2). We define the GEV model with a first order AR process for the state variable which we label GEV-AR model as

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

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

where $|\phi| < 1$. The state variable α_t is assumed to follow a stationary AR(1) process driven by 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 0 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}, $j = 1, \ldots, t-2$, weighted by

ϕ^j like for a standard AR model, however, the innovations arise from the Gumbel rather than the normal distribution. We 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, since 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, since the state equation is non-Gaussian, while the observation equation is non-linear. In the present paper, we use the idea of auxiliary mixture sampling and approximate the non-linear, non-Gaussian state

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

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

space model (5) and (6) by a very accurate finite mixture of non-linear 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 model 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 Type I extreme value (or 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:

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

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 $i = 1, \ldots, 10$ is reproduced in Table 1.

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 $s \equiv \{s_1, \ldots, s_{n-1}\}$, equations (5) and (6) form a non-linear Gaussian state space model, where:

$$\begin{aligned} \alpha_{t+1} &= m_{s_t} + \phi\alpha_t + v_{s_t} u_t, \qquad u_t \sim N(0,1), \qquad t = 1, \ldots, n-1, & (9) \\ y_t &= \mu + \psi\frac{\exp(\xi\alpha_t) - 1}{\xi} + e_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \qquad t = 1, \ldots, n, & (10) \end{aligned}$$

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

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 inverted Gamma posterior $\sigma^2 \,|\, \mu, \psi, \xi, \alpha, y$ is straightforward, while sampling from $(\mu, \psi, \xi) \,|\, \sigma^2, \alpha, y$ 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 non-linear 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 non-linear 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) 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 j-th sample from the approximated model, for $j = 1, \dots, M$, 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

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

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 approximates the target density well. 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 $Y_t = \{y_j\}_{j=1}^t$, and 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. However, 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 our empirical studies 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 $y_{t+1} \approx \mu + \psi\{\exp(\xi\alpha_{t+1}) - 1\}/\xi$. Then 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 $y_+ = \max(y, 0)$. Since 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

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

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

1. Initialize $t = 1$, 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 $\bar{w}_1 = \frac{1}{I}\sum_{i=1}^I w_1^i, \bar{W}_1 = \frac{1}{I}\sum_{i=1}^I W_1^i$.
   (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

$$\begin{aligned} w_t^i &= \frac{f(y_{t+1}|\alpha_{t+1}^i, \omega) f(\alpha_{t+1}^i | \alpha_t^i, \omega) \hat{f}(\alpha_t^i | Y_t, \omega)}{g(\alpha_{t+1}^i, \alpha_t^i | Y_{t+1}, \omega)} = \frac{f(y_{t+1}|\alpha_{t+1}^i, \omega) \hat{f}(\alpha_{t+1}^i | \alpha_t^i, \omega)}{g(\alpha_{t+1}^i | y_{t+1}, \omega)}, \\ W_t^i &= \frac{F(y_{t+1}|\alpha_{t+1}^i, \omega) f(\alpha_{t+1}^i | \alpha_t^i, \omega)}{g(\alpha_{t+1}^i | y_{t+1}, \omega)}, \quad i = 1, \ldots, I, \end{aligned}$$

   and save $\bar{w}_t = \frac{1}{I}\sum_{i=1}^I w_t^i$, $\bar{W}_t = \frac{1}{I}\sum_{i=1}^I W_t^i$.
   (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 $\sum_{t=1}^n \log \bar{w}_t \xrightarrow{p} \sum_{t=1}^n \log f(y_t | Y_{t-1}, \omega)$.

In Section 5.5, we show that our proposed filter outperforms other filters, where we also consider the modified APF (MPF) similar to the filter proposed by Chamú Morales (2005) in the context of moving maxima processes. It implements an APF for the usual observation y_{t+1} less than a threshold, say, 95th percentile of y_t's. But, for the extreme observation y_{t+1} which exceeds the threshold, it uses the importance function based on a mixture distribution, $q(\alpha_{t+1}|y_{t+1}, \alpha_t, \omega) = 0.95 g(\alpha_{t+1}|\alpha_{t+1} > u_{t+1}, \omega) + 0.05 f(\alpha_{t+1}|\alpha_t, \omega)$, where $g(\alpha_{t+1}|\alpha_{t+1} > u_{t+1}, \omega)$ is the truncated Gumbel density with 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.

2.4 Extension to a threshold model

Since the GEV model is intended to describe the distribution of extreme observations, it is sometimes applied only to those extremes which exceed a high threshold (see e.g., Coles (2001)). As an extension of the time-dependent GEV model, we consider the threshold model only for the extremes that exceed a certain fixed value. Let δ denote the threshold and $\tilde{y}_t$ be a censored state variable. We assume that an extreme value, y_t, is observed only when $\tilde{y}_t \geq \delta$ holds. In the GEV-AR model, we modify the observation equation in the following way:

$$
\begin{aligned}
y_t &= \begin{cases} \tilde{y}_t, & \text{if } \tilde{y}_t \geq \delta, \\ \text{N.A.}, & \text{if } \tilde{y}_t < \delta, \end{cases} \\
\tilde{y}_t &= \mu + \psi \frac{\exp(\xi\alpha_t) - 1}{\xi} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \quad t = 1, \ldots, n.
\end{aligned}
$$

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

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

where $TN_{(-\infty,\,\delta)}$ denotes a 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 deal with the uncertainty of choosing the threshold δ within a Bayesian inference (e.g., Tancredi et al. (2006)). However, since our main focus is capturing the time-dependence in an extreme value process, we put this issue aside and assume that δ has been 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

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

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. As before, the model reduces to iid observations from the GEV distribution, if both θ and σ^2 are zero.

The initial state α_1 is assumed to be

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

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

3.2 Bayesian Estimation

The unknown model parameters of the GEV-MA model are equal to $\omega \equiv (\lambda, \sigma^2, \theta)$. We pursue a 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 Subsection 2.2. Finally, 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 $g(\eta_t)$ of the Gumbel distribution underlying the innovations η_t in 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 $t = 1, \dots, n-1$:

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

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

where $u_0 \sim N(0,1)$. Evidently, conditional on s, equations (5), (15) and (16) form a non-linear Gaussian state space model.

To perform MCMC estimation, we introduce 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 in the GEV-MA model, $\pi(\omega, u, s|y)$, we implement the following MCMC algorithm.

Algorithm 2: MCMC algorithm for the GEV-MA model

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

Steps 1 to 3 are implemented as in Subsection 2.2. Note, however, that we do not marginalize over s, when sampling the parameter θ. Also sampling the latent mixture indicator variables s in Step 4 is different, because they are no longer independent given 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}. Hence, s_t affects not only the distribution of y_{t+1} as in Subsection 2.2, but also the distribution of y_{t+2}, and 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 $\pi(s_t|\omega, u_{t-1}, u_t, u_{t+1}, u_{t+2}, s_{t-1}, s_{t+2}, y_{t+1}, y_{t+2})$.

Finally, we sample in Step 5 the disturbance u to obtain α through the state equations (15) and (16). Once more, we apply the block sampler developed by Omori and Watanabe (2008) for non-linear Gaussian state space models to sample u. 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: $\mu \sim N(0,\ 10)$, $\psi \sim$ Gamma(2, 2), $\xi \sim N(0,\ 4)$, $\sigma^2 \sim$ IG(2.5, 0.025), $(\phi+1)/2 \sim$ Beta(4, 4), $(\theta+1)/2 \sim$ Beta(4, 4). We draw M = 20,000 samples after the initial 10,000 samples are discarded as the burn-in period. The computational results are generated using Ox version 4.02 (Doornik (2006)).

(i) GEV-AR model

Parameter	True	Mean	Stdev.	95% interval	Inefficiency
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

Parameter	True	Mean	Stdev.	95% interval	Inefficiency
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 2: Estimation result of the GEV models for simulated data.

Table 2 gives 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 $b_w = 1{,}000$.

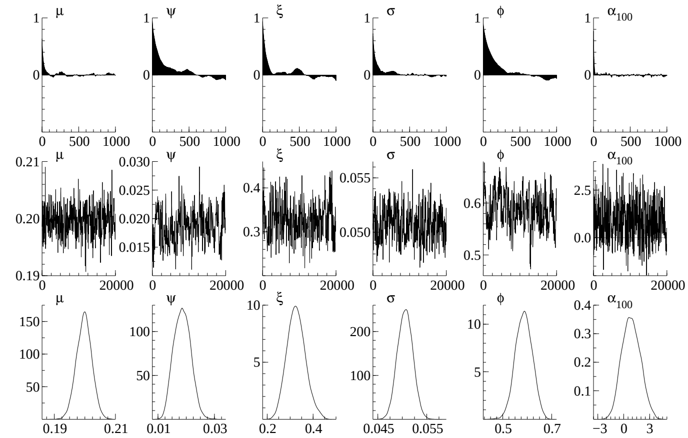

Figure 1: GEV-AR model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

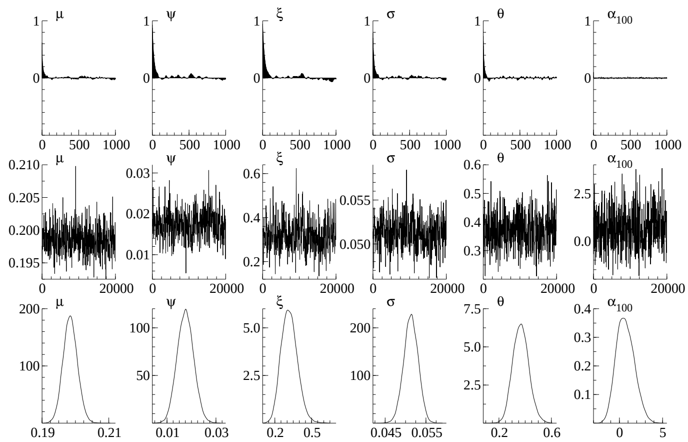

Figure 2: GEV-MA model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

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.

Figures 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.

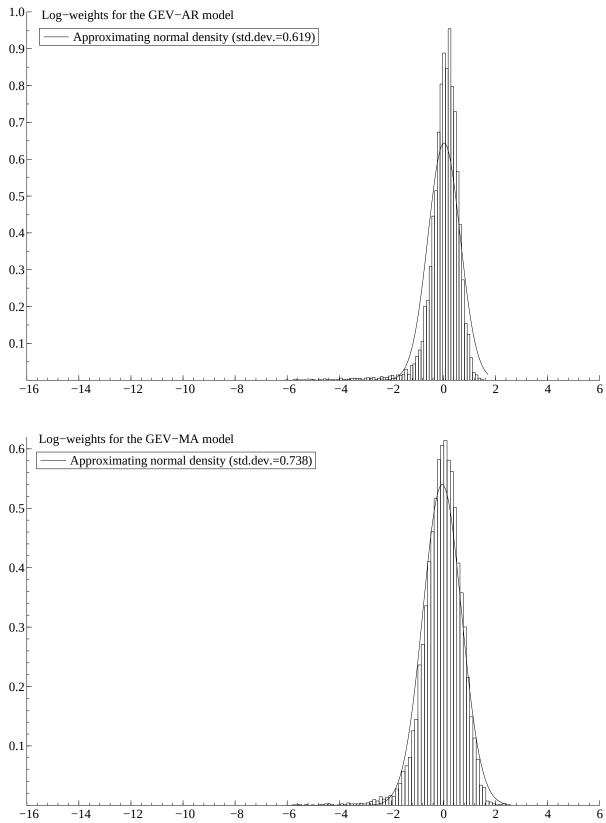

Figure 3: Histogram of the $\log(w_j \times M)$ for M = 20,000 iterations for simulated data. GEV-AR (top) and GEV-MA (bottom) models.

When ξ is close to zero, there could be some confounding between ξ and σ in sampling procedure, since $\exp(\xi\alpha_t) \approx 1 + \xi\alpha_t + \xi^2/2\alpha_t^2$ and $y_t \approx \mu + \psi\alpha_t + \psi\xi/2\alpha_t^2 + \varepsilon_t$. We checked this point using simulated data with $\xi = 10^{-2}$ and 10^{-3}, but the scatter plot of ξ and σ exhibited no problematic strong correlations.

As described in Subsection 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.

Figure 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 given in (8).

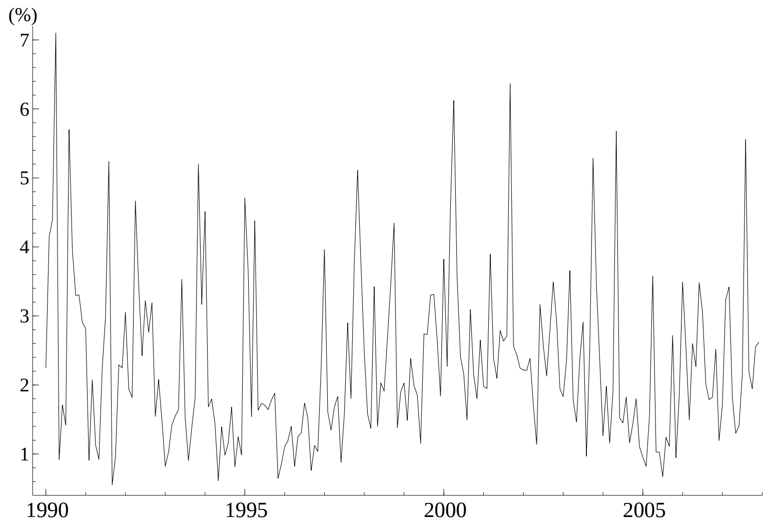

Figure 4: Minimum return data for the TOPIX (multiplied by -1, 1990/Jan – 2007/Dec).

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). While there would be many frequencies to analyze financial market variables, the daily stock return is one of the most popular figures that market participants much care about. Moreover their extreme values, especially in the left tail of the distribution, on monthly basis are of interest for a wide range of applications in financial econometrics.

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 Figure 4 shows the corresponding time series. The skewness is positive and the kurtosis is larger than that of a normal distribution, which implies a longer right tail and fatter tails.

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

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

5.2 Estimation results

We 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.

Table 4 reports the estimation result of the various 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 interval of the corresponding parameter does not contain zero. From these results we find evidence of time-dependence in the minimum returns and find both autoregressive and moving average effects in the GEV-AR and the GEV-MA model, respectively.

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

Parameter	GEV	GEV-AR	GEV-MA
μ	1.6987 (0.0680)	1.5817 (0.0763)	1.5495 (0.0789)
	[1.5709, 1.8371]	[1.4312, 1.7329]	[1.3949, 1.7018]
	216.8	190.0	141.6
ψ	0.8042 (0.0546)	0.7308 (0.0728)	0.6912 (0.0746)
	[0.7005, 0.9149]	[0.5668, 0.8609]	[0.5197, 0.8249]
	226.1	186.7	181.6
ξ	0.1519 (0.0714)	0.2115 (0.0647)	0.2212 (0.0629)
	[0.0328, 0.3237]	[0.0938, 0.3470]	[0.1100, 0.3557]
	422.8	185.3	171.2
σ	0.1052 (0.0324)	0.1443 (0.0648)	0.1714 (0.0754)
	[0.0605, 0.1849]	[0.0657, 0.3334]	[0.0678, 0.3436]
	66.7	197.3	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)	2.0018 (0.2186)	2.0851 (0.2433)
	[1.4929, 2.2971]	[1.1659, 2.4905]	[1.6242, 2.5965]
	338.3	161.3	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: Estimation result of the GEV models for the TOPIX minima data.

the rescaled maximum is H_ξ, then 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 distributions, as often pointed out in the financial literature.

Figures 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.

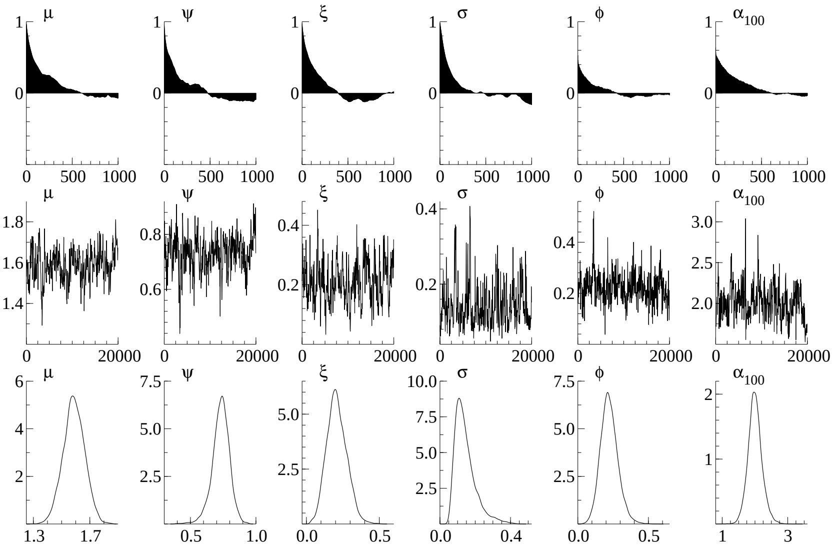

Figure 5: Estimation result of the GEV-AR model for the TOPIX minima data.

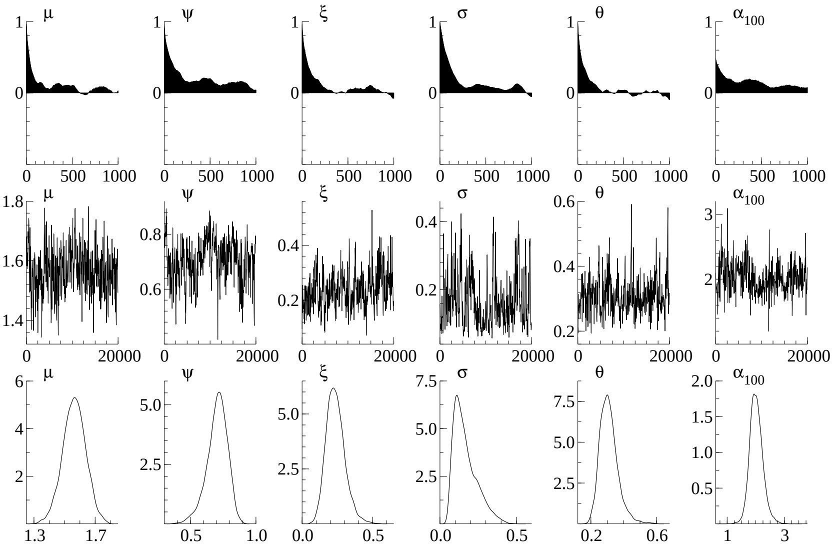

Figure 6: Estimation result of the GEV-MA model for the TOPIX minima data.

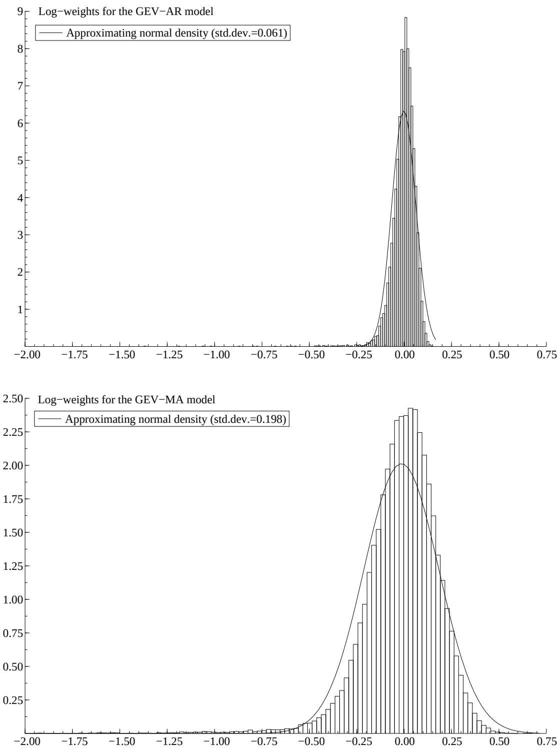

Figure 7: Histogram of the $\log(w_j \times M)$ for $M = 20{,}000$ iterations for TOPIX minima data. GEV-AR (top) and GEV-MA (bottom) models.

Figure 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 very small variance, indicating that our approximation based on the mixture of normals is very accurate.

5.3 Posterior predictive analysis

To check the plausibility of our proposed model, we conducted a posterior predictive analysis (see, e.g., (Gelman et al., 2003, Chapter 6)). We generated a set of $n = 216$ new observations for each MCMC draw and calculated for each data set twelve summary statistics, namely 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 2–3.

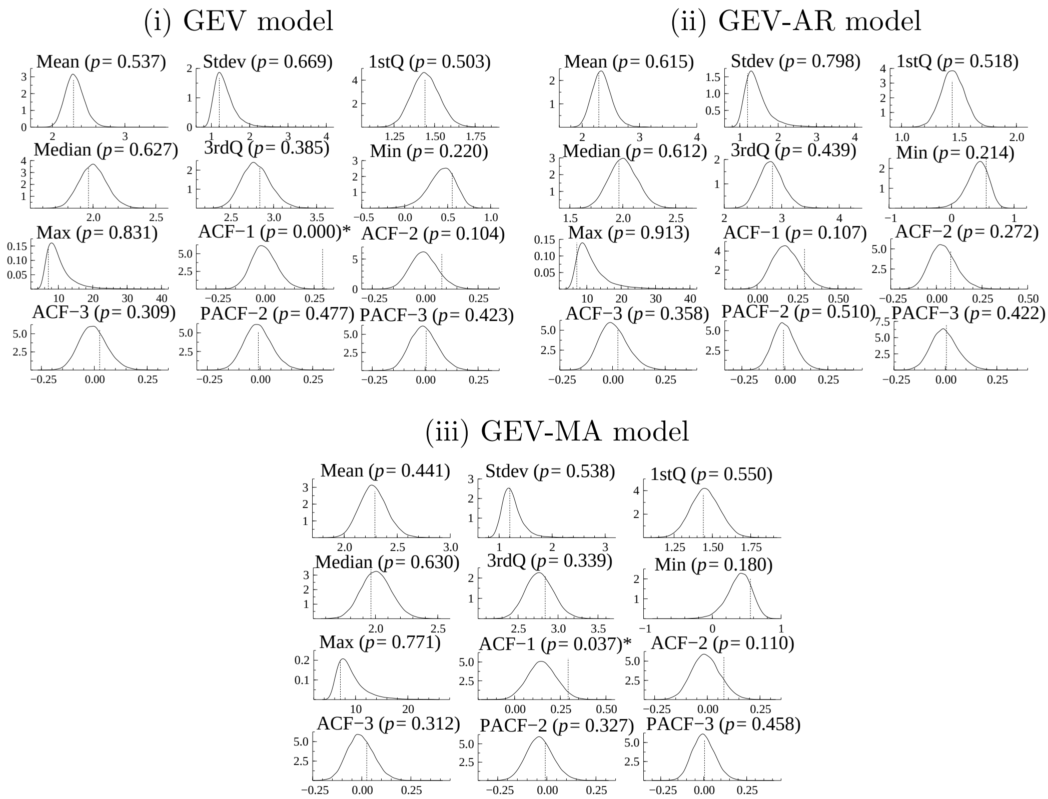

Figure 8: Posterior predictive check. The vertical lines denote the values of test quantities based on the observed data, and the posterior predictive p-value are in parentheses. (*) indicates a statistical significance at 10% significant level.

Using the posterior predictive distributions of these statistics, we are able to 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.

Figure 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 substantial first-order autocorrelation.

5.4 Model comparison using marginal likelihoods

This section conducts a model comparison 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.

Thus, using marginal likelihoods, we compare the three models of the time-dependent GEV class and the threshold models introduced in Section 2.4. For the threshold models we set the threshold δ equal to 1.48 (the 90th percentile of all daily returns multiplied by -1) and use only the minima (multiplied by -1) that exceed the threshold. The original observations under the threshold are treated as censored variables in that model.

	(i) Standard models			(ii) Threshold models		
Model	GEV	GEV-AR	GEV-MA	GEV	GEV-AR	GEV-MA
Likelihood ordinate	-312.23	-304.39	-306.21	-317.64	-305.95	-308.59
(S.E.)	(0.26)	(0.32)	(0.28)	(0.11)	(0.57)	(0.79)
Prior ordinate	-1.86	-3.00	-4.06	-2.15	-2.52	-2.50
Posterior ordinate	9.64	10.70	8.39	9.25	10.00	10.48
(S.E.)	(0.29)	(0.47)	(0.50)	(0.63)	(0.75)	(0.30)
Marginal likelihood	-323.75	-318.11	-318.68	-329.05	-318.47	-321.58
(S.E.)	(0.39)	(0.57)	(0.58)	(0.64)	(0.94)	(0.85)

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

Table 5: Estimated marginal likelihood for the TOPIX minima data.

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. The likelihood function is computed by the particle filters developed in Section 5.5 using $I = 10,000$ particles. Ten replications of the filter are implemented to obtain the standard error of the likelihood.

Table 5 reports the results of marginal likelihood estimation. We refer to the GEV models without the threshold as the standard models. The likelihood ordinates of the GEV-AR and GEV-MA models are larger than that of the GEV model. Although the marginal likelihoods of these models are penalised by the prior or the posterior ordinates due to additional parameters, they imply that the time-dependent GEV models still outperform the GEV model.

The marginal likelihoods for the threshold models are smaller than those for the standard models overall, but, among them, the time-dependent models still outperform the GEV model. Taking account of the standard errors of the marginal likelihoods, the GEV-AR model is found to be the best model among the threshold models for the TOPIX returns minima data.

5.5 Comparison of particle filters

In addition, we compare the particle filters discussed in Section 2.3 using the number of particles $I = 10000, 50000, 100000$ and the number of replications $M = 10, 100$.

In Table 6, for GEV-AR and GEV-MA models, it is clear that our new filter (New) produces more stable and accurate estimates than the auxiliary particle filter (APF) and the modified APF (MPF). The APF leads to quite unstable estimates and does not work in the existence of extreme observations. The simple particle filter (PF) also produces stable estimates, but their standard errors are much larger than those of our filter.

For the GEV model, estimates are stable for all filters , but the standard errors of our filter are found to be much smaller than those of MPF and PF. Overall the estimation results support that our method outperforms the PF, MPF and APF.

5.6 Forecasting performance

In this section, we investigate the forecasting performance of the three competing models. Consider the one-step ahead predictive density given data y, $\pi(y_{n+1}|y)$,

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

	GEV-AR				GEV-MA			
	New	PF	MPF	APF	New	PF	MPF	APF
$R = 10,000$	-304.39	-304.41	-310.66	-341.49	-306.21	-306.63	-321.81	-367.26
$M = 10$	(0.32)	(0.40)	(2.67)	(3.72)	(0.28)	(0.32)	(2.80)	(5.70)
$R = 50,000$	-304.28	-304.36	-306.85	-330.23	-306.30	-306.33	-316.61	-351.83
$M = 10$	(0.09)	(0.16)	(1.66)	(4.99)	(0.11)	(0.24)	(2.83)	(4.80)
$R = 100,000$	-304.33	-304.50	-305.67	-327.34	-306.27	-306.41	-312.55	-348.08
$M = 10$	(0.08)	(0.27)	(1.92)	(3.90)	(0.05)	(0.20)	(2.26)	(3.26)
$R = 100,000$	-304.35	-304.37	-306.41	-327.39	-306.29	-306.32	-313.07	-346.20
$M = 100$	(0.09)	(0.18)	(1.76)	(4.37)	(0.08)	(0.18)	(2.53)	(4.63)

	GEV		
	New	PF	MPF
$R = 10,000$	-312.23	-312.20	-312.47
$M = 10$	(0.26)	(0.83)	(0.51)
$R = 50,000$	-312.01	-312.14	-312.29
$M = 10$	(0.12)	(0.18)	(0.23)
$R = 100,000$	-312.10	-312.12	-312.17
$M = 10$	(0.07)	(0.19)	(0.11)
$R = 100,000$	-312.14	-312.12	-312.14
$M = 100$	(0.11)	(0.20)	(0.15)

*All values are in natural log scale. Standard errors are in parentheses. R and M denote the number of particles and iterations respectively.

Table 6: 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).

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 i-th iteration given the current sample of parameters and latent variables $(\omega^i, s^i, \alpha^i)$.

To compare the time-dependent extreme value models, we compute the mean and median of the one-step ahead predictive distribution for fifty observations given a fixed number of past observations, namely 165, in our stock returns data set. Specifically, we first use the sample period from January 1990 to September 2003 giving a total of 165 observations to estimate parameters using MCMC and generate samples from the one-step ahead predictive distribution for the minimum daily stock return in October 2003. Next we use the sample period from February 1990 to October 2003 to predict a dependent variable in November 2003. We repeat this rolling estimation until we obtain fifty one-step ahead predicted means and medians. The number of the MCMC iterations and the prior settings are the same as those in the preceding estimations.

Table 7 reports the mean absolute percentage errors (MAPE) and root mean squared percentage errors (RMSPE), namely,

$$\mathrm{MAPE} \;=\; \frac{1}{50}\sum_{i=1}^{50}\left|\frac{\hat{y}_{N+i}-y_{N+i}}{y_{N+i}}\right|, \quad \mathrm{RMSPE} = \left\{\frac{1}{50}\sum_{i=1}^{50}\left(\frac{\hat{y}_{N+i}-y_{N+i}}{y_{N+i}}\right)^2\right\}^{1/2},$$

where $N = 165$, which is the sample size of each subsample, and y_{N+i}, $\hat{y}_{N+i}$ denote the actual value and the posterior predictive value (mean or median) at period $N+i$ using the subsample data from period i to $N+i-1$.

	(i) Predictive mean			(ii) Predictive median		
Model	GEV	GEV-AR	GEV-MA	GEV	GEV-AR	GEV-MA
MAPE	0.537	0.505	0.495	0.436	0.397	0.406
RMSPE	0.726	0.674	0.676	0.580	0.522	0.539

Table 7: Mean absolute percentage errors (MAPE) and root mean squared percentage errors (RMSPE) for the posterior estimates of predictive distribution.

The posterior predictive medians provide better forecasts overall, and the time-dependent GEV models obviously perform better than the simple GEV model. While the GEV model forecasts at an average level of the extreme values based on the historical data, the time-dependent models put an emphasis on the recent activity of the extreme values, which would yield better forecasts in the experiments. Among the competing GEV models, the posterior predictive median of the GEV-AR model performs better than others based on these measures.

We also computed the root mean squared errors (RMSE) for both the posterior predictive mean and median of predictive distribution and found that the posterior predictive median yield better forecasts, and the GEV-AR model outperforms the other models. However, since the RMSE's are heavily influenced by two huge values in $\{y_t\}_{t=166}^{215}$ which are forecasted, we focused on the MAPE and RMSPE to evaluate the forecasting performance.

6 Conclusion

In the context of extreme value modeling, this paper develops a new approach to model time-dependence in the GEV distribution using a state space representation where the state variables either follows an AR or an MA processes with innovations from the Gumbel distribution. Approximating the Gumbel density by a ten-component mixture of normal distributions, a mixture sampler is proposed to implement Markov chain Monte Carlo methods. A simulation study shows that the proposed estimation schemes produce draws from the posterior distribution of the time-dependent GEV models quite efficiently.

In our application, several competing models in the time-dependent GEV class including threshold models are fitted to the monthly series of minimum daily returns for the TOPIX data. The parameter estimates show that the TOPIX minima data exhibit time-dependence. Model comparison based on marginal likelihoods indicates that the time-dependent GEV models outperform the simple GEV model and, moreover, that the GEV-AR model provides the best fit to the TOPIX minima data. In addition, from a forecasting perspective, predictive distributions are estimated for fifty sub-sample periods. The results of the forecasting performance confirm that time-dependent GEV models, especially, the GEV-AR model, outperform the simple GEV model.

In the present paper we assumed for all time-dependent GEV models for simplicity that the idiosyncratic error ε_t appearing in the observation equation (5) follows a normal distribution. Since we are modeling extreme values, one might expect that the distribution of the idiosyncratic error is non-normal, skewed or fat-tailed. Our algorithm can be extended easily to deal with non-normal error disturbance in the observation equation. For instance, the extension to a skew-t error distribution is straightforward by adding a couple of steps to our sampling algorithm.

However, the posterior predictive analysis performed in Subsection 5.3 for the TOPIX minima data indicated that a model based on a normal error distribution seems to be plausible in the sense that it replicates the characteristics of the observations. Thus we leave such distributional extensions for future work.

A 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 $R = \{\psi : \psi \leq 0\}$, $\lambda^* \sim TN_R(\lambda_*, \Sigma_*)$, where

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

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 $\sigma^2|\lambda, \alpha, y \sim \text{IG}(\hat{n}/2, \hat{S}/2)$, 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, $(\alpha_{k_{i-1}}, \ldots, \alpha_{k_i - 1})$ for $i = 1, \ldots, K+1$ with $k_0 = 0$ and $k_{K+1} = n$, we use the stochastic knots given by $k_i = \text{int}\,[n(i + U_i)/(K+2)]$, 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 non-linear state space model

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

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

$$\begin{aligned}\pi(u_{r-1},\ldots,u_{r+d-1}|\vartheta) &\propto \prod_{t=r}^{r+d}\exp\left[-\frac{1}{2\sigma^2}\left\{y_t-\mu-\psi\frac{\exp(\xi\alpha_t)-1}{\xi}\right\}^2\right] \\ &\times \prod_{t=r-1}^{r+d-1}\exp\left(-\frac{u_t^2}{2}\right)\times f(\alpha_{r+d}), \qquad (17)\end{aligned}$$

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)

$$h(\alpha_t) \equiv -\frac{1}{2\sigma^2}\left\{y_t-\mu-\psi\frac{\exp(\xi\alpha_t)-1}{\xi}\right\}^2, \qquad (18)$$

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,

$$\begin{aligned}h(\alpha_t) &\approx h(\hat{\alpha}_t)+h'(\hat{\alpha}_t)(\alpha_t-\hat{\alpha}_t)+\frac{1}{2}h''(\hat{\alpha}_t)(\alpha_t-\hat{\alpha}_t)^2 \\ &= \frac{1}{2}h''(\hat{\alpha}_t)\left\{\alpha_t-\left(\hat{\alpha}_t-\frac{h'(\hat{\alpha}_t)}{h''(\hat{\alpha}_t)}\right)\right\}^2+\text{const.} \\ &= -\frac{(y_t^*-\alpha_t)^2}{2\sigma_t^{*2}}+\text{const.}, \qquad (19)\end{aligned}$$

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

$$
\begin{aligned}
\sigma^{*2}_{r+d} &= \frac{1}{-h''(\hat{\alpha}_{r+d}) + \phi^2/v^2_{s_{r+d}}}, \\
y^*_{r+d} &= \sigma^{*2}_{r+d} \left\{ h'(\hat{\alpha}_{r+d}) - h''(\hat{\alpha}_{r+d})\hat{\alpha}_{r+d} + \phi(\alpha_{r+d+1} - m_{s_{r+d}})/v^2_{s_{r+d}} \right\},
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
y^*_t &= \alpha_t + \sigma^*_t \zeta_t, \quad t = r, \ldots, r+d, \\
\alpha_{t+1} &= m_{s_t} + \phi\alpha_t + v_{s_t} u_t, \quad t = r-1, \ldots, r+d-1, \\
(\zeta_t,\, u_t)' &\sim N(0,\, I_2), \quad t = r, \ldots, r+d.
\end{aligned} \tag{20}
$$

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 $(y^*_r, \ldots, y^*_{r+d})$, $(\sigma^{2*}_r, \ldots, \sigma^{2*}_{r+d})$ in (20) (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 $(y^*_r, \ldots, y^*_{r+d})$, $(\sigma^{2*}_r, \ldots, \sigma^{2*}_{r+d})$.

3. Run Kalman filter and the disturbance smoother (e.g. Koopman (1993)) using the current points $(y^*_r, \ldots, y^*_{r+d})$, $(\sigma^{2*}_r, \ldots, \sigma^{2*}_{r+d})$ in (20) 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. Since 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

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

where

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

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

$$
\begin{aligned}
\pi(u_{r-1}, \ldots, u_{r+d-1}|\vartheta) \;\propto\; & \prod_{t=r}^{r+d} \exp\left[-\frac{1}{2\sigma^2} \left\{ y_t - \mu - \psi \frac{\exp(\xi z \gamma_t) - 1}{\xi} \right\}^2 \right] \\
& \times \prod_{t=r-1}^{r+d-1} \exp\left(-\frac{u_t^2}{2} \right) \times f(y_{r+d+1}|\omega, \alpha_{r+d+1}),
\end{aligned}
\tag{21}
$$

where

$$
\begin{aligned}
& f(y_{r+d+1}|\omega, \alpha_{r+d+1}) \\
&= \begin{cases} \exp\left[-\frac{1}{2\sigma^2} \left\{ y_{r+d+1} - \mu - \psi \frac{\exp(\xi \alpha_{r+d+1}) - 1}{\xi} \right\}^2 \right], & \text{if } r + d < n, \\ 1, & \text{if } r + d = n, \end{cases}
\end{aligned}
\tag{22}
$$

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 (22) 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. PhD 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 J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, pp. 3569–3649. Amsterdam: North-Holland.

Chib, S. and I. Jeliazkov (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.* London: Springer.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis 13*, 257–272.

Doornik, J. (2006). *Ox: Object Oriented Matrix Programming.* London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (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.* New York: Springer.

Frühwirth-Schnatter, S. and R. Frühwirth (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis 51*, 3509–3528.

Frühwirth-Schnatter, S., R. Frühwirth, L. Held, and H. Rue (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing 19*, forthcoming.

Frühwirth-Schnatter, S. and H. Wagner (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. and M. Grigoletto (2004). Smoothing sample extremes with dynamic models. *Extremes 7*, 221–236.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference* (2 ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis* (2nd ed.). London: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, Volume 4, pp. 169–188. New York: Oxford University Press.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley.

Huerta, G. and B. Sansó (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics 14*, 285–299.

Hughes, G. L., S. S. Rao, and T. S. Rao (2007). Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. *Proceedings of the Royal Society A 463*, 241–259.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Koop, G. (2003). *Bayesian Econometrics*. Chichester: Wiley.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*, 117–126.

Leadbetter, K. R., G. Lindgren, and H. Rootzén (2004). *Extremes and related properties of random sequences and series*. New York.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics 140*, 425–449.

Omori, Y. and T. Watanabe (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis 52*, 2892–2910.

Pitt, M. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Shephard, N. and M. Pitt (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 B. Finkenstadt and H. Rootzen (Eds.), *Extreme Values in Finance, Telecommunications and the Environment*, Chapter 1, pp. 1–78. London: Chapman and Hall/CRC Press.

Smith, R. L. and J. E. Miller (1986). A non-Gaussian state space model and application to prediction of records. *Journal of Royal Statistical Society 48*, 79–88.

Smith, R. L. and I. Weissman (1996). Characterization and estimation of the multivariate extremal index. Manuscript.

Tancredi, A., C. Anderson, and A. O'Hagan (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes 9*, 87–106.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). *Biometrika 91*, 246–248.

Zhang, Z. and R. L. Smith (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability 41*, 1113–1123.