

CIRJE-F-952

**Bayesian Modeling of Dynamic Extreme Values:
Extension of Generalized Extreme Value
Distributions with Latent Stochastic Processes**

Jouchi Nakajima
Bank of Japan

Tsuyoshi Kuniyama
Duke University

Yasuhiro Omori
The University of Tokyo

January 2015

CIRJE Discussion Papers can be downloaded without charge from:

<http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html>

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Bayesian modeling of dynamic extreme values: Extension of generalized extreme value distributions with latent stochastic processes

Jouchi Nakajima^{a,*}, Tsuyoshi Kuniyama^b, Yasuhiro Omori^c

^a*Bank of Japan, Tokyo, 103-8660, Japan.*

^b*Department of Statistical Science, Duke University, Durham 27708-0251, U.S.A.*

^c*Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan.*

Abstract

This paper develops Bayesian inference of extreme value models with a flexible time-dependent latent structure. The generalized extreme value distribution is utilized to incorporate state variables that follow an autoregressive moving average (ARMA) process with Gumbel-distributed innovations. The time-dependent extreme value distribution is combined with heavy-tailed error terms. An efficient Markov chain Monte Carlo algorithm is proposed using a state space representation with a mixture of normal distribution approximating the Gumbel distribution. The methodology is illustrated using extreme data of stock returns and electricity demand. Estimation results show the usefulness of the proposed model and evidence that the latent autoregressive process and heavy-tailed errors plays an important role to describe the monthly series of minimum stock returns and maximum electricity demand.

Keywords: ARMA process, Extreme values, Generalized extreme value distribution, Markov chain Monte Carlo, Mixture sampler, Stock returns, Electricity demand.

*Corresponding author. Tel.: +81 3 3279 1111; fax: +81 3 5201 6525. The views expressed herein are those of the authors alone and do not necessarily reflect those of the Bank of Japan.

Email addresses: `jouchi.nakajima@boj.or.jp` (Jouchi Nakajima), `tsuyoshi.kuniyama@stat.duke.edu` (Tsuyoshi Kuniyama), `omori@e.u-tokyo.ac.jp` (Yasuhiro Omori)

1. Introduction

Extreme value models have been widely studied for analysis of excessively large or small rare events in application to environmental sciences, engineering and financial econometrics. Much progress has been made in the general area of extreme value models, such as generalized extreme value (GEV) distribution and generalized Pareto distribution, to address extremal behavior of random variables (e.g., Coles (2001)). Exploring statistical property of the maxima and minima is practically of importance because an extremely rare event typically leads to huge damage or loss for human life such as earthquake and financial crisis.

Recently, a number of articles have shed light on time-dependent or clustering structure of extremes in time series context. One of standard and pragmatic approaches is to introduce autoregressive (AR) process or local time trend for extreme value model parameters (e.g., Smith and Miller (1986), Gaetan and Grigoletto (2004), Huerta and Sansó (2007), Sang and Gelfand (2009), Liu et al. (2012)). Zhao et al. (2011) utilize an idea of generalized autoregressive conditional heteroscedasticity (GARCH) to a scale parameter in the GEV model. From another perspective, in the context of max-stable process (Smith (2003)), the moving maxima process has been extended in a dynamic manner to the maxima of moving maxima process (e.g., Deheuvels (1983), Smith and Weissman (1996), Zhang and Smith (2004), Chamú Morales (2005), Kuniyama et al. (2012)).

Among them, Nakajima et al. (2012) extend the GEV distribution to incorporate the time-dependence using a state space representation where state variables either follow a stationary AR process or moving average (MA) process with innovations arising from a Gumbel distribution. This dynamic extreme value model is a natural extension associated with a Box-Cox transformation where a response of observation equation follows the GEV distribution, apart from the standard normal-distributed error models. Nakajima et al. (2012) provide an empirical evidence that the time-dependent extreme value models yield plausible estimation results for financial data and perform better than the standard GEV model. Following their analysis, the current paper examines two important extensions for the the Gumbel-type time-dependent GEV model: (i) the latent variable follows a stationary autoregressive moving average (ARMA) process; and (ii) the observational error follows Student's t -distribution. This considerably flexible extreme value model is designed in order to more generally and practically address a wide variety of time-dependence for time series of extremes. An efficient Markov

chain Monte Carlo (MCMC) method is developed to compute parameter estimation and forecasting of the proposed model. The usefulness of the model and methodology is illustrated in analysis of stock returns as well as of electricity demand with emphasis on a strong evidence of time-dependence among the observed monthly minimum series of daily stock returns and monthly maximum series of hourly electricity demand.

The proposed time-dependent extreme value model has a great potential to be applied in a wide range of practical analysis. From one perspective in analysis of financial market, there is a pressing need for accurate forecasts of extremes of product prices in financial market from experience of the recent financial crisis. Risk management is nowadays at center of discussion both for market participants and monetary authorities (see e.g., McAleer et al. (2013)). Practically, volatility clustering possibly exists not only for high-frequency time series of, for example, stock return and foreign exchange rate return, but also its block maxima and minima. During market turbulent of financial crisis, it is reasonable to expect higher extremes arising from larger uncertainty market condition than other time periods. From another perspective in analysis of electricity industry, understanding of electricity demand is critical to managing sufficient supply capacity. Exceedance of electricity demand over the maximum supply capacity can severely damage our living life and economy. To avoid it and pursue an efficient investment strategy of electricity infrastructure, a reasonable time series model for maximum electricity demand is required, but little has been investigated for this point (e.g., Engle et al. (1992)). For both analyses, if the time-dependence structure exists in time series of interest, the proposed dynamic extreme value model would provide a better approximation to the underlying process and therefore yield more accurate forecasts than a standard static extreme value model.

The rest of paper is organized as follows. Section 2 develops the proposed time-dependent GEV model. Section 3 proposes the efficient MCMC algorithm and associated particle filter method to compute the likelihood function for model evaluation. Section 4 illustrates the methodology using simulated data. Section 5 provides the application to monthly minimum return series of daily stock data. Section 6 reports the application to monthly maximum series of hourly electricity demand. Finally, Section 7 concludes.

2. GEV-ARMA t model

2.1. Preliminary

To introduce the proposed model, we build on a standard GEV distribution given by

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

where $\psi > 0$, and $1 + \xi(y_t - \mu)/\psi > 0$. The Gumbel distribution is defined as a subset of the GEV distribution with the limit $\xi \rightarrow 0$. We here assume $\mu = 0$ and $\psi = 1$, which leads to the distribution of the standard Gumbel random variable:

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

We note that $E(\alpha_t) = c_0$, and $\text{Var}(\alpha_t) = \pi^2/6$ ($\equiv c_1$), where c_0 is the Euler constant.

A key idea to link the GEV distribution and latent process is to consider Y_t follows the GEV distribution of (1), and to define a latent process

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

which implies that α_t follows the Gumbel distribution $G(\cdot)$ of eqn. (2). This results in the following relation:

$$Y_t = \mu + \psi \frac{\exp(\xi \alpha_t) - 1}{\xi}, \quad (3)$$

where $\alpha_t \sim G$. Nakajima et al. (2012) assume parametric time series process such as AR and MA models for $\{\alpha_t\}$ where the innovation of the process follows the Gumbel distribution, and apply eqn. (3) as a measurement equation with an additional idiosyncratic shock following a normal distribution. In this paper, we generalize this dynamic extreme value model by assuming the stationary ARMA process for α_t and the Student's t -distribution for the idiosyncratic error in the measurement equation.

2.2. The model

Suppose a time series of extremes, $y = \{y_1, \dots, y_n\}$, follows the model

$$y_t = \mu + \psi \frac{\exp(\xi \alpha_t) - 1}{\xi} + \sqrt{\lambda_t} \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad t = 1, \dots, n, \quad (4)$$

$$\alpha_{t+1} = \phi \alpha_t + \eta_t + \theta \eta_{t-1}, \quad \eta_t \sim G, \quad t = 1, \dots, n-1, \quad (5)$$

where $|\phi| < 1$ and $|\theta| < 1$. The state variable α_t is assumed to follow the stationary and invertible ARMA(1, 1) process driven by the innovations $\{\eta_t\}_{t=0}^n$ that follow the Gumbel distribution defined by eqn. (2), where $(\eta_t, \eta_{s \neq t})$ are mutually independent. We specify the heavy-tail measurement error in eqn. (4); $\sqrt{\lambda_t} \varepsilon_t$ is assumed to follow the Student's t -distribution with unknown degrees of freedom ν by letting $\lambda_t^{-1} \sim \text{Gamma}(\nu/2, \nu/2)$. For the initial state α_1 , we assume a normal distribution with mean and variance equal to those of the stationary distribution of the ARMA process (5); namely, $\alpha_1 \sim N(a_0^* c_0 + \eta_0, a_1^* c_1)$, where $a_0^* = (\theta + \phi)/(1 - \phi)$, and $a_1^* = (\theta + \phi)^2/(1 - \phi^2)$.

The parameters (ϕ, θ) measure time-dependence of the latent process that plays an important role to dynamically predict y_t on the basis of the GEV distribution. A salient feature of our model is that the response y_t would be a sequence of independently and identically distributed from the GEV distribution in the limiting case $\phi = \theta = \sigma = 0$.

We refer to the model of eqns. (4)-(5) as GEV-ARMA t model. Note that $\lambda_t \equiv 1$ for all t implies a normal-error measurement equation, referred as GEV-ARMA model. The GEV-ARMA model reduces to GEV-AR (if $\theta \equiv 0$) or GEV-MA (if $\phi \equiv 0$) model of Nakajima et al. (2012).

3. Estimation

3.1. Bayesian inference

The eqns. (4)-(5) form a non-linear and non-Gaussian state space model, therefore, implementation of maximum likelihood estimation is computationally intensive. To overcome it, we employ a Bayesian inference for the GEV-ARMA t model (see e.g., MacDonald (2011) for review of Bayesian inference for various extreme value models). Assuming prior distributions for the model parameters, a posterior distribution of the model is evaluated based on sample generation via Markov chain Monte Carlo (MCMC) computations. To exploit traditional, efficient MCMC methods for Gaussian state space models (e.g., Prado and West (2010)) for the GEV-ARMA t model, a

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).

strategy taken here is that we approximate the non-Gaussian innovation by a highly accurate finite mixture of normal density. Some works on data augmentation approach with latent variables provide a fast and efficient estimation scheme as described in the following subsections.

3.2. Mixture sampler

Approximation and auxiliary sampling with a finite normal mixture, so called a *mixture sampler*, have become a standard toolkit to tackle non-Gaussian models (see e.g., Frühwirth-Schnatter (2006)). An early and popular example is a stochastic volatility model where the $\log \chi_1^2$ density is approximated by a seven- or ten-component normal mixture (Kim et al. (1998); Omori et al. (2007)). Our estimation method for the GEV-ARMA model relies on the mixture sampler for the Gumbel distribution developed by Frühwirth-Schnatter and Frühwirth (2007), Frühwirth-Schnatter and Wagner (2006), and Frühwirth-Schnatter et al. (2009).

Let $g(\cdot)$ denote the probability density function of the Gumbel distribution of eqn. (2); namely, $g(\eta_t) = \exp(-\eta_t - e^{-\eta_t})$. This density is approximated by a normal mixture of $K = 10$ components:

$$g(\eta_t) \approx \sum_{i=1}^K p_i f_N(\eta_t | m_i, v_i^2),$$

where $f_N(\eta_t | m_i, v_i^2)$ denotes the normal density with mean m_i and variance v_i^2 , for $i = 1, \dots, K$. Frühwirth-Schnatter and Frühwirth (2007) provide the selection of (p_i, m_i, v_i^2) for $K = 10$ (see Table 1).

We employ a latent mixture indicator variable, $s_t \in \{1, \dots, K\}$, for $t = 1, \dots, n-1$, which refers to an index of the mixture component at each time point with a prior probability $\Pr(s_t = i) = p_i$. A typical missing data approach applies to the conditionally normal density, $\pi(\eta_t | s_t = i) = f_N(\eta_t | m_i, v_i^2)$. Formally, conditional on $s \equiv (s_1, \dots, s_{n-1})$ and $\lambda = (\lambda_1, \dots, \lambda_n)$, eqns. (4) and (5) lead to the form of a Gaussian state space model:

$$y_t = \mu + \psi \frac{\exp(\xi \alpha_t) - 1}{\xi} + \sqrt{\lambda_t} \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad (6)$$

$$\alpha_{t+1} = \phi \alpha_t + (m_{s_t} + v_{s_t} u_t) + \theta (m_{s_{t-1}} + v_{s_{t-1}} u_{t-1}), \quad u_t \sim N(0, 1), \quad (7)$$

$$\alpha_1 = a_0^* c_0 + (m_{s_0} + v_{s_0} u_0^*) + \sqrt{a_1^* c_1} \delta_0^*, \quad u_0^*, \delta_0^* \sim N(0, 1). \quad (8)$$

An efficient filtering and smoothing sampler on non-linear Gaussian state space models is now applicable to generate $u = (u_0, u_1, \dots, u_{n-1})$, where $u_0 \equiv (\delta_0^*, u_0^*)'$; the MCMC method for simulation of the full joint posterior $\pi(\omega, \nu, s, u, \lambda | y)$, where $\omega = (\mu, \psi, \xi, \sigma, \phi, \theta)$, is constructed in the next subsection.

3.3. MCMC algorithm

We assume traditional priors for the parameters (ω, ν) : normal distributions for μ and ξ , gamma distributions for ψ and ν , an inverse gamma distribution for σ^2 , and shifted beta distributions for ϕ and θ ; these priors are all mutually independent among the parameters in (ω, ν) . We outline components of the MCMC computations here, and provide details in Appendix A.

The MCMC algorithm for the GEV-ARMA model is implemented as follows:

1. Initialize ω, ν, s, u , and λ .
2. Generate $(\mu, \psi, \xi) | \sigma, \phi, \theta, s, u, \lambda, y$.
3. Generate $\sigma | \mu, \psi, \xi, \phi, \theta, s, u, \lambda, y$.
4. Generate $(\phi, \theta) | \mu, \psi, \xi, \sigma, s, u, \lambda, y$.
5. Generate $s | \omega, u, \lambda, y$.
6. Generate $u | \omega, s, \lambda, y$.
7. Generate $\lambda | \omega, \nu, s, u, y$.
8. Generate $\nu | \lambda$.
9. Go to 2.

The generation of the parameters (ω, ν) and λ uses a direct sampling or standard Metropolis-Hasting (MH) sampling strategy. For sampling u in Step 6, we employ a

block (multi-move) sampler developed by Omori and Watanabe (2008) for non-linear Gaussian state space models. This sampler divides a sequence of the state vector (u_0, \dots, u_{n-1}) into several blocks and draws sample from the conditional posterior distribution of each block given its adjacent blocks; the block sampler is known to produce more efficient draws than a single-move sampler that generates one state u_t at a time given others $\{u_s; s \neq t\}$ (e.g., Shephard and Pitt (1997)). The generation of s in Step 5 only requires sampling from K -point discrete distribution at each time point. Our experiences with the MCMC algorithm summarized here are that mixing is good and no pitfall is observed in practice. Simulation example illustrates performance of the sampler in Section 4.

3.4. Likelihood evaluation

Following Nakajima et al. (2012), we evaluate the likelihood of the proposed model using an efficient auxiliary particle filter method (APF, Pitt and Shephard (1999)). Define $f(y|\Theta)$ as the likelihood function for a fixed model parameter $\Theta \equiv (\omega, \nu)$, and let $f(y_t|\alpha_t, \Theta)$ denote the measurement density implied by (4), and $f(\alpha_{t+1}|\alpha_t, \Theta)$ implied by (5) denote the state evolution density from (5). The particle filtering is basically based on

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

where $Y_t = \{y_j\}_{j=1}^t$. The particles are simulated from $f(\alpha_t | Y_t, \Theta)$ to produce a discrete uniform approximation $\hat{f}(\alpha_t | Y_t, \Theta)$ to $f(\alpha_t | Y_t, \Theta)$. This approach follows a simple and standard particle filter and may yield almost zero importance weights for the extreme observations (see e.g., Chamú Morales (2005)). To avoid this issue, the APF employed in this paper is implemented as follows.

We consider the approximation of $y_{t+1} \approx \mu + \psi \{\exp(\xi \alpha_{t+1}) - 1\} / \xi$, ignoring the disturbance. We write

$$m_{t+1} \equiv \frac{1}{\xi} \log \left(1 + \xi \frac{y_{t+1} - \mu}{\psi} \right)_+ \approx \alpha_{t+1}, \quad t = 1, \dots, n-1, \quad (9)$$

where $y_+ = \max(y, 0)$. As pointed out by Nakajima et al. (2012), m_{t+1} can be considered as the most likely value of the state α_{t+1} given y_{t+1} . Then, we use the importance

function

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

where this importance density generates α_{t+1} from the Gumbel distribution with the mode m_{t+1} . The APF is implemented with the following procedure. We first generate $\alpha_1^i \sim N(c_0/(1 - \phi), c_1/(1 - \phi^2))$, for $i = 1 : I$, where $i = 1 : I$ denotes $i = 1, \dots, I$. We then 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 $\bar{W}_1 = \frac{1}{I} \sum_{i=1}^I W_1^i$. We set $\hat{f}(\alpha_1^i | y_1, \Theta) = w_1^i / \sum_{j=1}^I w_1^j$, $i = 1 : I$. For $t = 1, \dots$, we iterate the following steps:

1. Generate $(\alpha_{t+1}^i, \alpha_t^i)$, $i = 1 : I$, from the importance function $g(\alpha_{t+1}, \alpha_t | Y_{t+1}, \Theta)$.
2. Compute

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

and save $\bar{w}_t = \frac{1}{I} \sum_{i=1}^I w_t^i$, and $\bar{W}_t = \frac{1}{I} \sum_{i=1}^I W_t^i$.

3. Set $\hat{f}(\alpha_{t+1}^i | Y_{t+1}, \Theta) = w_t^i / \sum_{j=1}^I w_t^j$, $i = 1 : I$.

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

4. Simulation analysis

4.1. Setup

We illustrate the estimation method developed in the previous section using simulated data. A series of $n = 1,000$ observations is generated from the GEV-ARMA model with fixed parameter values $\mu = 0.1$, $\psi = 0.02$, $\xi = 0.3$, $\sigma = 0.1$, $\phi = 0.5$, $\theta = 0.3$, and $\nu = 15$. In the following analyses, we assume the priors: $\mu \sim N(0, 10)$, $\psi \sim \text{Gamma}(2, 2)$, $\xi \sim N(0, 1)$, $\sigma^2 \sim \text{IG}(2.5, 0.025)$, $(\phi + 1)/2 \sim \text{Beta}(4, 4)$, $(\theta + 1)/2 \sim \text{Beta}(4, 4)$, $\nu \sim \text{Gamma}(16, 0.8)$.

For the estimation of the MCMC algorithm, we draw $M = 100,000$ samples after the initial 50,000 samples are discarded as the burn-in period. To check efficiency of the sampling algorithm, we compute the inefficiency factor (see e.g., Chib (2001)), defined

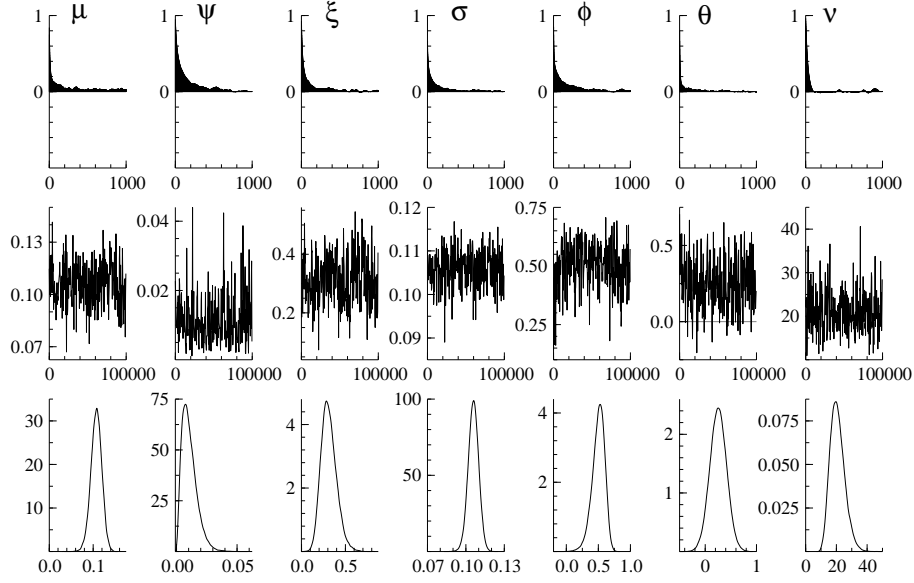


Figure 1: Estimation results of the GEV-ARMA model for simulated data. Sample autocorrelations (top), sample paths (middle) and posterior densities (bottom).

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; 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 (e.g., 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 $b_w = 1,000$.

Parameter	True	Mean	Stdev.	95% interval	Ineff.
μ	0.1	0.1070	0.0124	[0.0818, 0.1304]	70.1
ψ	0.02	0.0107	0.0063	[0.0020, 0.0258]	147.6
ξ	0.3	0.3113	0.0880	[0.1603, 0.5038]	85.8
σ	0.1	0.1059	0.0041	[0.0977, 0.1137]	68.1
ϕ	0.5	0.4997	0.0997	[0.2780, 0.6660]	95.3
θ	0.3	0.2545	0.1628	[-0.0652, 0.5747]	36.4
ν	15	20.730	4.7236	[12.769, 31.119]	50.1

Table 2: Estimation results of the GEV-ARMA model for simulated data. Ineff refers to the inefficiency factors.

4.2. Estimation result

Table 2 reports posterior estimates and the inefficiency factors for the parameters and a selected state variable, α_{100} . It shows that all posterior means are close to the true values, and that the true values are contained in the 95% credible intervals. The inefficiency factors are comparable to those in the previous work (Nakajima et al. (2012)). Figure 1 shows the sample autocorrelation functions, the sample paths and the posterior densities for the simulation analysis. The sample paths look stable and the sample autocorrelations drop smoothly, which confirms that while the proposed model forms the considerably flexible structure, the proposed estimation scheme provides plausible posterior estimates with reasonable efficiency.

5. Application to extremes of stock returns

5.1. Data

This section applies the proposed model to extremes of stock returns. We focus on monthly minima of daily and 5-minute intraday returns. We consider detecting time-varying behavior of the minima, relative to maxima, is more important from a viewpoint of risk management that market participants much care about. Specifically, We use daily and 5-minute intraday returns of the Tokyo Stock Price Index (TOPIX) and pick up the monthly minima. The minima are multiplied by -1 for estimation.

The data span from January 1980 to December 2009 ($n = 360$) for the daily return, and from April 1996 to March 2005 ($n = 108$) for the intraday return. Figure 2 plots the time series, which shows the former series includes two large (in absolute value) minima caused by the Black Monday in October 1987 and by the Lehman shock in October 2008. To model these stock return extremes we expect the time-dependent structure and heavy-tail assumption with Student t -distribution incorporated into the GEV model. Table 3 reports descriptive statistics of the minima data. It is clear that the (unconditional) skewness is positive and the kurtosis is larger than that of a normal distribution, which shows the fatter tails with the longer-right tail.

Series	Observations	Mean	Stdev.	Skewness	Kurtosis	Max.	Min.
Daily-based	360	2.095	1.527	3.052	22.21	15.81	0.122
Intraday-based	108	1.395	1.153	2.682	11.40	6.751	0.220

Table 3: Summary statistics for the TOPIX minima data (multiplied by -1).

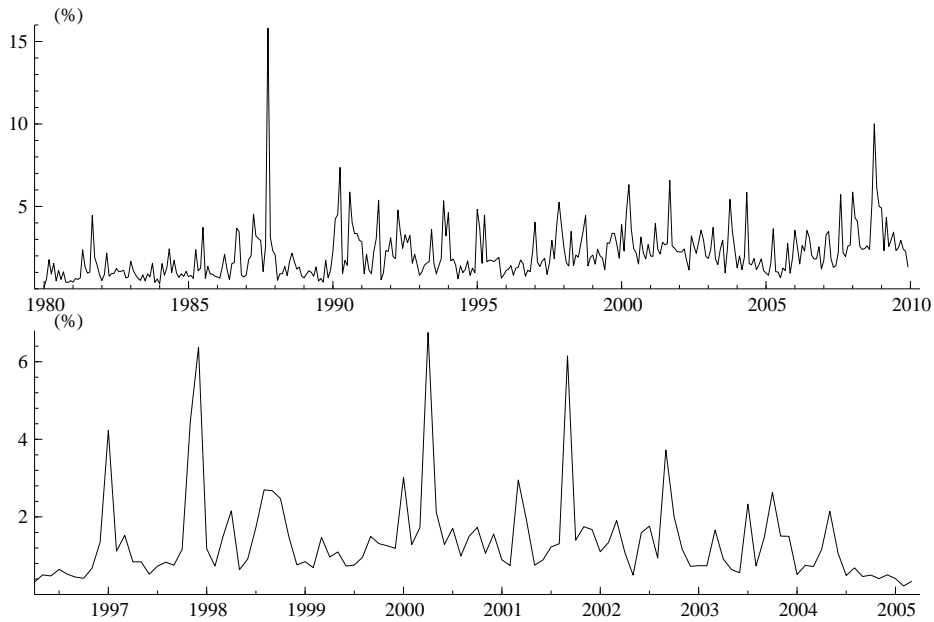


Figure 2: Time series of TOPIX minimum data (multiplied by -1).

5.2. Model comparisons

We compare models in the class of the GEV-ARMA t based on the marginal likelihoods using the TOPIX minima data. The competing models are summarized as follows:

1. Normal-error models ($\lambda_t \equiv 1$)

- **GEV**: no time-dependence ($\phi = \theta = 0$);
- **GEV-AR**: AR time-dependence ($\theta = 0$);
- **GEV-MA**: MA time-dependence ($\phi = 0$);
- **GEV-ARMA**: ARMA time-dependence.

2. Student's t -error models

- **GEV-t**: no time-dependence ($\phi = \theta = 0$);
- **GEV-ARt**: AR time-dependence ($\theta = 0$);
- **GEV-MAt**: MA time-dependence ($\phi = 0$);
- **GEV-ARMA t** : ARMA time-dependence.

This analysis covers various time-dependent structures, namely, AR(1), MA(1), and ARMA(1, 1) processes, in the class of GEV-ARMA_t. The estimation method for the GEV-ARMA_t model proposed in this paper is applicable to all the other models listed above, while the algorithm reduces to the simpler form in the case of the GEV-AR model (see Nakajima et al. (2012)).

The marginal likelihood is a standard measure to compare models in Bayesian analysis. It is defined as the integral of the likelihood function with respect to the prior density of the parameters. With the prior probabilities of the competing models assumed to be equal, we select the model which yields the largest marginal likelihood. In our analysis, the marginal likelihood is estimated using the method of Chib (1995). Specifically, the log of marginal likelihood is given by

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

where Θ is the parameters in the model, $m(y)$ is the marginal likelihood, $f(y|\Theta)$ is the likelihood function, $\pi(\Theta)$ is the prior probability density, and $\pi(\Theta|y)$ is the posterior density. The likelihood function is computed by the particle filter using $I = 10,000$ particles. To evaluate the posterior density $\pi(\Theta|y)$, we use the method of Chib (1995) and Chib and Jeliazkov (2001) with 10,000 draws obtained through reduced MCMC runs. Although the above equality holds for any Θ , we use the posterior mean of Θ to obtain a stable estimate of $m(y)$.

Table 4 reports estimated marginal likelihoods of eight competing models for the TOPIX minima data. This result shows three main findings. First the marginal likelihood of the GEV/GEV- t model is lower than the other time-dependent GEV models, which indicates that incorporating the time-dependence contributes to the model fit of the GEV model. Second, the marginal likelihoods of the GEV-AR/AR_t model is higher than those of the GEV-MA/MAt model, which implies that the autoregressive process fits to the TOPIX minima data than the moving average process embedded in the dynamic GEV model. These results are consistent with those of Nakajima et al. (2012). In the results here, it is interesting that the marginal likelihood of the GEV-ARMA/ARMA_t model is lower than the GEV-AR/AR_t and GEV-MA/MAt models. The TOPIX minima data prefers the simpler autoregressive process than the ARMA, for the underlying structure of the GEV model.

Third, the marginal likelihoods of the model with the Student's t -errors are higher than those with the normal errors, which implies the heavier-tailed distribution is

(i) Daily-return-based data (1980–2009)					
Model	Log-ML	(S.E.)	Model	Log-ML	(S.E.)
GEV	-564.91	(0.73)	GEV-t	-549.67	(0.80)
GEV-AR	-513.55	(0.67)	GEV-ARt	-498.32	(0.67)
GEV-MA	-531.38	(0.69)	GEV-MAt	-517.62	(0.59)
GEV-ARMA	-557.78	(0.78)	GEV-ARMAt	-543.22	(0.82)

(ii) Intraday-return-based data (1996–2005)					
Model	Log-ML	(S.E.)	Model	Log-ML	(S.E.)
GEV	-127.17	(0.62)	GEV-t	-123.73	(0.40)
GEV-AR	-110.67	(0.69)	GEV-ARt	-107.38	(0.65)
GEV-MA	-117.50	(0.51)	GEV-MAt	-114.06	(0.49)
GEV-ARMA	-124.72	(1.34)	GEV-ARMAt	-115.80	(0.44)

Table 4: Estimated marginal likelihoods (ML) for the TOPIX minima data. All values are in natural log scale. Standard errors (S.E.) are in parentheses.

better to describe the idiosyncratic errors than the normal in each time-dependent GEV model for both data series. Estimated posterior means of the state variable α_t is mostly lower than that of the GEV-AR model, because relatively larger observations are described by the heavy-tailed idiosyncratic error rather than by the state variable, which is preferred by the TOPIX minima data. As a result, the estimated marginal likelihoods indicate the GEV-ARt model is the best model for the TOPIX minima data among the eight candidate models.

We provide an additional model comparison as a robustness check for findings shown above. Cotter (2007) proposes a use of GARCH filtered returns of financial variables for adjusting unconditional extreme values in an analysis of conditional tail estimation. In line with this idea, we fit GARCH(p, q) model to daily-based TOPIX minima data for $p = 1, 2$ and $q = 1, 2$, and select the lags based on the AIC. It results in the GARCH(1, 2) model, and filtered residuals (standardized by estimated volatility) are used for the model comparison for the time-dependent GEV models. Table 5 reports the estimated marginal likelihoods for the candidate models, which show the same order of model performance and still indicate the superiority of the GEV-ARt model.

5.3. Parameter estimates

We discuss posterior estimates of the model parameters, focusing on the time-dependent GEV models with the Student's t -errors, because the estimated marginal likelihoods clearly favors them compared to those with normal errors. Table 6 reports the posterior estimates of the GEV-ARt, GEV-MAt, and GEV-ARMAt models, which

GARCH-filtered daily-return-based data (1980–2009)

Model	Log-ML	(S.E.)	Model	Log-ML	(S.E.)
GEV	-426.61	(0.56)	GEV-t	-411.14	(0.71)
GEV-AR	-372.48	(0.59)	GEV-ARt	-357.15	(0.64)
GEV-MA	-392.57	(0.61)	GEV-MAt	-378.58	(0.41)
GEV-ARMA	-416.66	(1.11)	GEV-ARMAt	-408.46	(1.26)

Table 5: Estimated marginal likelihoods (ML) for the GARCH-filtered TOPIX minima data. All values are in natural log scale. Standard errors (S.E.) are in parentheses.

Parameter	GEV-ARt	GEV-MAt	GEV-ARMAt
μ	1.0764 (0.0834) [0.9066, 1.2350] 388.5	1.2335 (0.0582) [1.1180, 1.3491] 224.5	0.2486 (0.1749) [-0.0958, 0.5918] 141.7
ψ	0.5189 (0.0803) [0.3478, 0.6589] 475.5	0.6763 (0.0515) [0.5695, 0.7694] 271.0	0.0433 (0.0228) [0.0105, 0.0959] 478.1
ξ	0.3758 (0.0565) [0.2704, 0.4898] 385.4	0.2688 (0.0436) [0.1880, 0.3612] 438.1	0.2862 (0.0361) [0.2192, 0.3608] 264.4
σ	0.1897 (0.0580) [0.0892, 0.3094] 317.8	0.1343 (0.0473) [0.0671, 0.2444] 199.1	0.3583 (0.0559) [0.2563, 0.4741] 123.2
ϕ	0.5085 (0.0623) [0.3924, 0.6340] 333.0		0.9136 (0.0211) [0.8669, 0.9485] 294.5
θ		0.4117 (0.0403) [0.3414, 0.4981] 232.4	-0.5370 (0.0938) [-0.7009, -0.3370] 197.2
ν	19.314 (5.0047) [10.675, 30.200] 32.2	19.974 (4.9811) [11.492, 30.803] 26.8	18.956 (5.0060) [10.442, 30.000] 27.2

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 6: Estimation result of the time-dependent GEV models with heavy-tailed errors for the daily-based TOPIX minima data.

exhibits several variations among the models depending on their process structure. First, focusing on the posterior means of time-dependence parameters, ϕ in the GEV-ARt model is 0.5085, θ in the GEV-MAt model is 0.4117, and (ϕ, θ) in the GEV-ARMAAt model is (0.9136, -0.5370). The first autocorrelation of the underlying state variable implied by these estimates is roughly 0.5 in the GEV-ARt, 0.3 in the GEV-MAt, and 0.6 in the GEV-ARMAAt model. Based on the GEV-ARt, the best model in the model comparison above, 95% credible intervals do not include zero, which indicates the time-dependence exists in the state variable of the time-dependent GEV model fit to the minima series of stock returns.

With the variation in the estimates of the time-dependence parameters, the estimates of the GEV parameters (μ, ψ) and error variance σ are different among the models. The results show that posterior means of μ and ψ in the GEV-ARMAAt model are lower than those in the GEV-ARt and GEV-MAt models. Also, the parameter σ is estimated to be larger in the GEV-ARMAAt model relative to the other models. For the parameter ξ , the posterior means are positive and the 95% credible intervals do not contain zero regardless of the model specification. Regarding the degree of freedom, ν , associated with the Student's t -errors, the posterior mean is around 20 for the time-dependent GEV models. Figure 3 exhibits MCMC results for the GEV-ARMAAt model, the sample autocorrelations, sample paths, and posterior densities. These results shows that the Markov chains mix well and the proposed estimation scheme is feasible in practice.

6. Application to extremes of electricity demand

6.1. Data

This section provides empirical finding by applying the proposed method to extremes of electricity demand. We use data of hourly actual electricity demand, provided by Tokyo Electric Power Company (TEPCO). The data span from January 2008 to August 2014. The time series of hourly electricity demand exhibits typical seasonality. These are extracted by a simply regression with dummy variables of weekends, national holidays, and each week in a year, as well as a special event dummy for the Tohoku Region Pacific Coast Earthquake on March 3, 2011. We analyze monthly maxima of the seasonally adjusted hourly demand with the sample size $n = 80$.

Figure 4 plots the time series, which shows large hikes in 2008 and 2010. Table 7 reports descriptive statistics of the maxima. The skewness is positive and the kurtosis is slightly larger than that of a normal distribution.

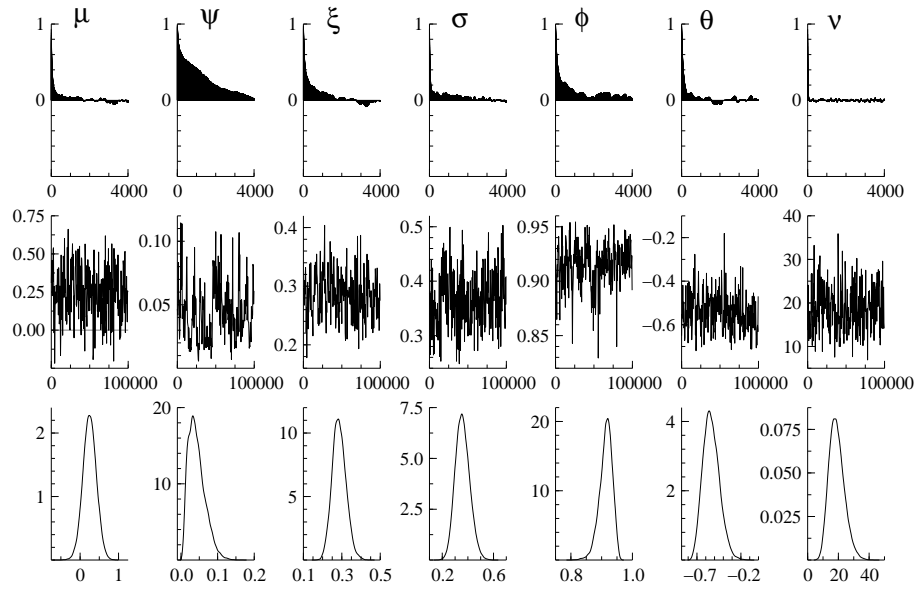


Figure 3: Estimation result of the GEV-ARMA model for the daily-based TOPIX minima data.

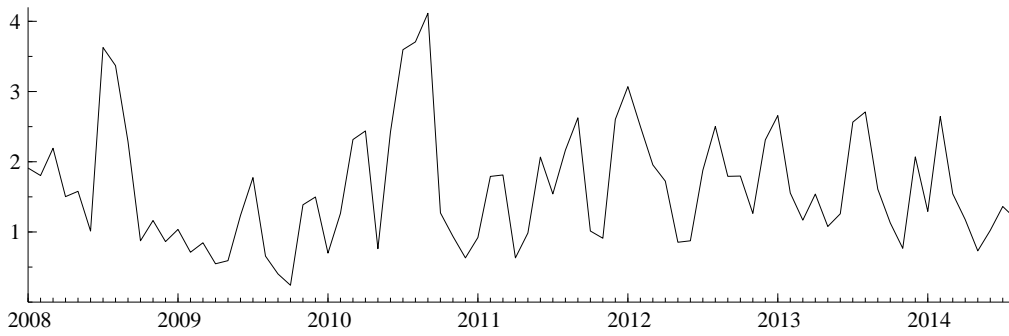


Figure 4: Time series of maximum electricity demand.

Observations	Mean	Stdev.	Skewness	Kurtosis	Max.	Min.
80	1.6238	0.8402	0.8462	3.2450	4.1142	0.2409

Table 7: Summary statistics for the maximum electricity demand data.

Model	Log-ML	(S.E.)	Model	Log-ML	(S.E.)
GEV	-102.03	(0.34)	GEV-t	-98.57	(0.31)
GEV-AR	-95.50	(0.45)	GEV-ARt	-92.61	(0.40)
GEV-MA	-95.00	(0.31)	GEV-MAt	-92.05	(0.49)
GEV-ARMA	-95.35	(0.40)	GEV-ARMAt	-92.41	(0.38)

Table 8: Estimated marginal likelihoods (ML) for the maximum electricity demand data. All values are in natural log scale. Standard errors (S.E.) are in parentheses.

6.2. Model comparison and parameter estimates

With the maximum electricity demand data, we estimate the competing models of the GEV-ARMA_t class in the same way as the previous section, and compute their marginal likelihoods for model comparison. Table 8 reports the estimated marginal likelihoods, which indicates that the marginal likelihoods of the time-dependent GEV models are higher than the one of the standard GEV model. The heavy-tailed models uniformly dominates the normal-error models. Among the time-dependent GEV models, the estimates of the marginal likelihood and standard errors imply the GEV-AR_t, GEV-MA_t, and GEV-ARMA_t are samely favored.

Table 9 reports the posterior estimates of the time-dependent GEV models with heavy-tailed errors. The estimates show little difference in the GEV parameters among the three models. Regarding the posterior means of time-dependence parameters, ϕ in the GEV-AR_t model is 0.3865, θ in the GEV-MA_t model is 0.4829, and (ϕ, θ) in the GEV-ARMA_t model is (0.2635, 0.2940). The credible intervals of the former two parameters does not include zero. These imply that the time-dependence exists in the underlying structure of the maximum electricity demand based on our proposed model. The degree of freedom ν in the Student's t -distribution is estimated around 20, which indicates moderate heavy-tails in the extreme series.

7. Conclusion

In this paper, we proposed the new elaborated time-dependent extreme value model, where the state variables follow the ARMA process, innovations follow the Gumbel distribution, and the idiosyncratic errors follow the Student's t -distribution. The advantage of this new model is the novel state space representation with the mixture of normal distribution approximating the Gumbel distribution, which enables us to develop the efficient MCMC algorithm to estimate the model. Illustration of the algorithm using simulated and empirical data shows reasonable performance of the estimation

Parameter	GEV-ARt	GEV-MAt	GEV-ARMAt
μ	1.0390 (0.1345)	1.0967 (0.1053)	1.0266 (0.1356)
	[0.7525, 1.2853]	[0.8882, 1.3024]	[0.7608, 1.2840]
	311.8	165.7	259.4
ψ	0.5372 (0.0773)	0.5322 (0.0738)	0.5228 (0.0791)
	[0.3887, 0.6942]	[0.3943, 0.6826]	[0.3721, 0.6845]
	196.4	118.1	179.7
ξ	0.1508 (0.1092)	0.1227 (0.0971)	0.1915 (0.1041)
	[-0.0418, 0.3848]	[-0.0533, 0.3182]	[-0.0016, 0.4078]
	151.0	148.8	152.9
σ	0.1097 (0.0360)	0.1153 (0.0428)	0.1095 (0.386)
	[0.0614, 0.2001]	[0.0617, 0.2280]	[0.0601, 0.2095]
	31.3	50.6	42.9
ϕ	0.3865 (0.0953)		0.2635 (0.1474)
	[0.2027, 0.5746]		[-0.0340, 0.5589]
	105.5		189.2
θ		0.4829 (0.0779)	0.2940 (0.1485)
		[0.3359, 0.6412]	[-0.0197, 0.5613]
		90.7	71.7
ν	20.003 (4.9571)	20.069 (4.9619)	20.012 (5.0176)
	[11.551, 30.998]	[11.538, 30.970]	[11.475, 30.958]
	9.4	7.8	5.7

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 9: Estimation result of the time-dependent GEV models with heavy-tailed errors for the maximum electricity demand data.

scheme. The empirical analysis using the TOPIX minima data obtained from daily and intraday returns indicates that the simple AR(1) process describes underlying dynamics in the TOPIX minima data based on the proposed time-dependent extreme value model better than independent or other processes such as MA(1) and ARMA(1, 1). Also, the estimation result implies that idiosyncratic shocks following the Student's t -distribution fits the data better than the normal distribution. The empirical results using the maximum electricity demand data implies that the time-dependent GEV models with the heavy-tailed errors outperform the standard and the time-dependent GEV models with the normal errors.

Acknowledgment

The authors would like to thank Sylvia Frühwirth-Schnatter for suggestion which motivated this paper. This work is supported by Nakajima Foundation, and the Grants-in-Aid for Scientific Research (A) 26245028 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. The computational results are generated using Ox version 6 (Doornik (2006)).

A. Appendix: MCMC algorithm

A.1. Generation of model parameters ω and ν

Let $\lambda = (\mu, \psi, \xi)'$ denote a vector of the GEV parameters. Sampling from the conditional posterior distribution $\pi(\lambda|\Theta_{-\lambda}, y)$, where $\Theta_{-\lambda} = (\sigma, \phi, \theta, s, u, \lambda)$, uses a MH algorithm with a proposal distribution based on a Taylor expansion method. We find the point $\hat{\lambda} = (\hat{\mu}, \hat{\psi}, \hat{\xi})'$ which (approximately) maximizes the conditional posterior. We generate a proposal, denoted by λ^* , from the truncated normal distribution, $\lambda^* \sim TN_R(\lambda_*, \Sigma_*)$, where $R = \{\psi : \psi \leq 0\}$,

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

Note that the mean and covariance are derived from a second-order Taylor expansion around the (approximate) mode λ^* . The MH algorithm accepts the candidate with probability

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

where λ denotes the current value, and $f_N(\cdot|\mu, \Sigma)$ denotes the probability density function of the multivariate normal distribution with mean μ and covariance matrix Σ .

The same strategy on the MH algorithm is employed to the generation of (ϕ, θ) and ν . Note that the truncation of the proposal distribution results in the support, $R = \{(\phi, \theta) : |\phi| < 1, |\theta| < 1\}$, and $R = \{\nu : \nu > 0\}$, respectively.

For the generation of σ , we assume a prior $\sigma^2 \sim IG(n_0/2, S_0)$, where IG denote an inverse gamma distribution. The generation of σ reduces to a direct sampling from

$IG(\hat{n}/2, \hat{S}/2)$, where

$$\hat{n} = n_0 + n, \quad \hat{S} = S_0 + \sum_{t=1}^n \frac{1}{\sqrt{\lambda_t}} \left\{ y_t - \mu - \psi \frac{\exp(\xi \alpha_t) - 1}{\xi} \right\}^2.$$

A.2. Generation of mixture variables s

The conditional posterior distribution of the mixture indicator variable s_t at time t involves with several adjacent state variables. Consider a transformation from (u_{t-2}, u_{t+1}) to (α_t, α_{t+3}) in the joint posterior probability density, which leads to the conditional posterior

$$\begin{aligned} & \pi(s_t | \omega, \lambda, u_{t-1}, u_t, u_{t+2}, \alpha_t, \alpha_{t+3}, s_{t-1}, s_{t+2}, y_{t+1}, y_{t+2}) \\ & \propto \sum_{j=1}^K \pi(s_t, s_{t+1} = j | \omega, \lambda, u_{t-1}, u_t, u_{t+2}, \alpha_t, \alpha_{t+3}, s_{t-1}, s_{t+2}, y_{t+1}, y_{t+2}), \end{aligned}$$

where

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

with

$$\begin{aligned} \alpha_{t+1} &= \phi \alpha_t + (m_{s_t} + v_{s_t} u_t) + \theta (m_{s_{t-1}} + v_{s_{t-1}} u_{t-1}), \\ \alpha_{t+2} &= \phi \alpha_{t+1} + (m_j + v_j \tilde{u}_{j,t+1}) + \theta (m_{s_t} + v_{s_t} u_t), \\ \tilde{u}_{j,t+1} &= \frac{1}{v_j(\phi + \theta)} [\alpha_{t+3} - \phi \{ \phi \alpha_{t+1} + m_j + \theta (m_{s_t} + v_{s_t} u_t) \} \\ & \quad - (m_{s_{t+2}} + v_{s_{t+2}} u_{t+2}) - \theta m_j]. \end{aligned}$$

Note that $\{\alpha_k\}_{k=t+3}^n$ are fixed because the resulting posterior distribution is conditional on $(\alpha_{t+3}, \{s_k\}_{k=t+2}^n)$. After the generation of s_t , we transform (α_t, α_{t+3}) back to (u_{t-2}, u_{t+1}) . For sampling s_0, s_{n-2} and s_{n-1} , the conditional posterior densities is computed without using $\tilde{u}_{j,t+1}$ defined above.

A.2.1. Generation of disturbance vector u

We implement the block (or multi-move) sampler which divides the vector of disturbance u into several blocks and samples each block given other blocks (Shephard and Pitt (1997), Watanabe and Omori (2004)). The sampler is built on the non-linear Gaussian state space form of the GEV-ARMA model:

$$y_t = \mu + \psi \frac{\exp(\xi z \gamma_t) - 1}{\xi} + \sqrt{\lambda_t} \sigma e_t, \quad z = (1, 0), \quad t = 1, \dots, n,$$

$$\gamma_{t+1} = w_t + T \gamma_t + H_t u_t, \quad \gamma_t = (\alpha_t, \beta_t)', \quad t = 0, \dots, n-1,$$

where

$$w_t = \begin{pmatrix} 1 \\ \theta \end{pmatrix} m_{s_t}, \quad T = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix}, \quad H_t = \begin{pmatrix} v_{s_t} \\ \theta v_{s_t} \end{pmatrix}, \quad t = 1, \dots, n-1,$$

$$w_0 = \begin{pmatrix} a_0^* c_0 + m_{s_0} \\ \theta m_{s_0} \end{pmatrix}, \quad H_0 = \begin{pmatrix} \sqrt{a_1^* c_1} & v_{s_0} \\ 0 & \theta v_{s_0} \end{pmatrix},$$

$$\begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim N(0, I_2), \quad t = 1, \dots, n, \quad u_0 = \begin{pmatrix} \delta_0^* \\ u_0^* \end{pmatrix} \sim N(0, I_2),$$

with $\gamma_0 = 0$. To sample a typical block, say, $(u_{r-1}, \dots, u_{r+d-1})$ from its joint conditional posterior distribution, we consider a transformation from (u_{r-3}, u_{r+d}) to $(\alpha_{r-1}, \alpha_{r+d+2})$ as in the generation of s_t . Define $\vartheta = (\omega, u_{r-2}, \alpha_{r-1}, u_{r+d}, \alpha_{r+d+2}, \{s_t\}_{t=r-1}^{r+d}, \{\lambda_t\}_{t=r-1}^{r+d}, \{\lambda_t, y_t\}_{t=r}^{r+d+1})$. Then the conditional posterior probability density is given by

$$\begin{aligned} & \pi(u_{r-1}, \dots, u_{r+d-1} | \vartheta) \\ & \propto \prod_{t=r}^{r+d} \exp \left[-\frac{1}{2\lambda_t \sigma^2} \left\{ y_t - \mu - \psi \frac{\exp(\xi \alpha_t) - 1}{\xi} \right\}^2 \right] \\ & \quad \times \prod_{t=r-1}^{r+d-1} \exp \left(-\frac{u_t^2}{2} \right) \times f(y_{r+d+1} | \omega, \alpha_{r+d+1}) f(\tilde{u}_{r+d}), \end{aligned} \quad (\text{A.1})$$

with

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

where $\{\alpha_t\}_{t=r}^{r+d+1}$ and \tilde{u}_{r+d} are computed from the state equations as follows:

$$\begin{aligned}
\alpha_{r+k} &= \phi\alpha_{r+k-1} + (m_{s_{r+k-1}} + v_{s_{r+k-1}}u_{r+k-1}) + \beta_{r+k-1}, \quad k = 0, \dots, d, \\
\beta_{r+k} &= \theta(m_{s_{r+k-1}} + v_{s_{r+k-1}}u_{r+k-1}), \quad k = 0, \dots, d, \\
\alpha_{r+d+1} &= \phi\alpha_{r+d} + (m_{s_{r+d}} + v_{s_{r+d}}\tilde{u}_{r+d}) + \beta_{r+d}, \\
\beta_{r+d+1} &= \theta(m_{s_{r+d}} + v_{s_{r+d}}\tilde{u}_{r+d}), \\
\tilde{u}_{r+d} &= \frac{1}{v_{s_{r+d}}(\phi + \theta)} \left\{ \alpha_{r+d+2} - \phi(\phi\alpha_{r+d} + m_{s_{r+d}} + \beta_{r+d}) \right. \\
&\quad \left. - (m_{s_{r+d+1}} + v_{s_{r+d+1}}u_{r+d+1}) - \theta m_{s_{r+d}} \right\}.
\end{aligned}$$

The posterior distribution is conditional on α_{r+d+2} so that the likelihood computation involves only with multiplication up to $t = r + d + 1$. We implement MH algorithm to sample from the conditional posterior (A.1).

A proposal distribution uses the approximation of the conditional posterior probability density; for $t = r, \dots, r + d - 1$ and $r + d = n$, we consider a Taylor expansion of the logarithm of the likelihood (excluding the constant term):

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

around the (approximated) mode, denoted by $\hat{\alpha}_t$. 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,

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

where $\sigma_t^{*2} = -\{h''(\hat{\alpha}_t)\}^{-1}$ and $y_t^* = \hat{\alpha}_t + \sigma_t^{*2}h'(\hat{\alpha}_t)$, for $t = r, \dots, r + d - 1$, and $t = r + d = n$. For $t = r + d < n$, we take a Taylor expansion of

$$\begin{aligned}
\tilde{h}(\gamma_{r+d}) &\equiv -\frac{1}{2\lambda_t\sigma^2} \left\{ y_{r+d} - \mu - \psi \frac{\exp(\xi\alpha_{r+d}) - 1}{\xi} \right\}^2 \\
&\quad - \frac{1}{2\lambda_t\sigma^2} \left\{ y_{r+d+1} - \mu - \psi \frac{\exp(\xi\alpha_{r+d+1}) - 1}{\xi} \right\}^2 - \frac{\tilde{u}_{r+d}^2}{2},
\end{aligned}$$

around the mode, denote by $\hat{\gamma}_{r+d}$, which yields

$$\tilde{h}(\gamma_{r+d}) \approx -\frac{1}{2}(y_{r+d}^* - \gamma_{r+d})' \Sigma_{r+d}^{*-1} (y_{r+d}^* - \gamma_{r+d}) + \text{const.},$$

where $\Sigma_{r+d}^* = -\left\{ \tilde{h}''(\hat{\gamma}_{r+d}) \right\}^{-1}$ and $y_{r+d}^* = \hat{\gamma}_{r+d} + \Sigma_{r+d}^* \tilde{h}'(\hat{\gamma}_{r+d})$, with $\tilde{h}'(\gamma_t) = d\tilde{h}(\gamma_t)/d\gamma_t$ and $\tilde{h}''(\gamma_t) = d^2\tilde{h}(\gamma_t)/d\gamma_t d\gamma_t'$. The proposal distribution given by

$$q(u_{r-1}, \dots, u_{r+d-1} | \vartheta) \propto \prod_{t=r}^{r+d-1} \exp \left\{ -\frac{(y_t^* - \alpha_t)^2}{2\sigma_t^{*2}} \right\} \times \prod_{t=r-1}^{r+d-1} \exp \left(-\frac{u_t^2}{2} \right) \\ \times \exp \left\{ -\frac{(y_{r+d}^* - \gamma_{r+d})' \Sigma_{r+d}^{*-1} (y_{r+d}^* - \gamma_{r+d})}{2} \right\}. \quad (\text{A.4})$$

Note that this is the posterior probability density of $(u_{r-1}, \dots, u_{r+d-1})$ for the state space model

$$\begin{aligned} y_t^* &= z\gamma_t + \sigma_t^* \zeta_t, & t &= r, \dots, r+d-1, \\ \gamma_{t+1} &= w_t + T\gamma_t + H_t u_t, & t &= r-1, \dots, r+d-1, \\ \begin{pmatrix} \zeta_t \\ u_t \end{pmatrix} &\sim N(0, I_2), & t &= r, \dots, r+d-1, \end{aligned} \quad (\text{A.5})$$

where

$$y_{r+d}^* = \begin{cases} z\gamma_{r+d} + \sigma_{r+d}^* \zeta_{r+d}, & \zeta_{r+d} \sim N(0, 1), & (\text{if } r+d = n) \\ \gamma_{r+d} + \Sigma_{r+d}^{*1/2} \zeta_{r+d}, & \zeta_{r+d} \sim N(0, I_2). & (\text{if } r+d < n) \end{cases}$$

A key strategy here is to draw a candidate $(u_{r-1}, \dots, u_{r+d-1})$ by exploiting Kalman filter and the simulation smoother (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) based on the linear Gaussian state space model (A.5). We compute the conditional modes $(\hat{\alpha}_r, \dots, \hat{\alpha}_{r+d})$ by repeating the following steps for several times until the smoothed state variables converge:

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

4. Replace $(\hat{\alpha}_r, \dots, \hat{\alpha}_{r+d})$ by $(\hat{\alpha}_r^*, \dots, \hat{\alpha}_{r+d}^*)$.
5. Go to 2.

Following Shephard and Pitt (1997), the blocks are stochastically determined every MCMC iteration to efficiently generate samples. Consider dividing the sequence of the state variables, (u_1, \dots, u_{n-1}) , into $K + 1$ blocks: $(u_{k_{i-1}+1}, \dots, \alpha_{k_i})$ for $i = 1, \dots, K + 1$ with $k_0 = 0$ and $k_{K+1} = n$. We use a so-called *stochastic knots* that generates the edges based on $k_i = \text{int}[n(i + U_i)/(K + 2)]$, for $i = 1, \dots, K$, where U_i is a random sample from a uniform distribution $U[0, 1]$.

A.3. Generation of scale variable λ

The conditional posterior distribution of λ reduces to a direct sampling of $\lambda_t | \omega, \nu, s, u, y \sim IG(\hat{\nu}/2, \hat{S}_t/2)$, where

$$\hat{\nu} = \nu + 1, \quad \hat{S}_t = \nu + \frac{1}{\sigma^2} \left\{ y_t - \mu - \psi \frac{\exp(\xi \alpha_t) - 1}{\xi} \right\}^2,$$

for $t = 1, \dots, n$.

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.
- Cotter, J. (2007). Varying the VaR for unconditional and conditional environments. *Journal of International Money and Finance* 26, 1338–1354.

- 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.
- Engle, R. F., C. Mustafa, and J. Rice (1992). Modelling peak electricity demand. *Journal of Forecasting* 11, 241–251.
- 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, 479–492.
- 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.
- 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.
- Huerta, G. and B. Sansó (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics* 14, 285–299.

- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* 80, 117–126.
- Kunihama, T., Y. Omori, and Z. Zhang (2012). Efficient estimation and particle filter for max-stable processes. *Journal of Time Series Analysis* 33, 61–80.
- Liu, Y., T. Bahadori, and H. Li (2012). Sparse-GEV: Sparse latent space model for multivariate extreme value time series modeling. Manuscript.
- MacDonald, A. E. (2011). Extreme value mixture modelling with medical and industrial applications. PhD dissertation, Department of Mathematics and Statistics, University of Canterbury.
- McAleer, M., J.-A. Jiménez-Martín, and T. Pérez-Amaral (2013). International evidence on GFC-robust forecasts for risk management under the Basel Accord. *Journal of Forecasting* 32, 267–288.
- Nakajima, J., T. Kunihama, Y. Omori, and S. Frühwirth-Schnatter (2012). Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Computational Statistics and Data Analysis* 56, 3241–3259.
- 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.
- Prado, R. and M. West (2010). *Time Series Modeling, Computation, and Inference*. New York: Chapman & Hall/CRC.
- Sang, H. and A. Gelfand (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* 16, 407–426.

- 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.
- 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.
- Zhao, X., C. J. Scarrott, L. Oxley, and M. Reale (2011). GARCH dependence in extreme value models with Bayesian inference. *Mathematics and Computers in Simulation* 81, 1430–1440.