Bayesian Estimation and Particle Filter for Max-Stable Processes

Tsuyoshi Kunihama
Graduate School of Economics, University of Tokyo

Yasuhiro Omori
University of Tokyo

Zhengjun Zhang
University of Wisconsin Madison

August 2010

Bayesian estimation and particle filter for max-stable processes

TSUYOSHI KUNIHAMA

Graduate School of Economics, University of Tokyo, Tokyo 113-0033, Japan

YASUHIRO OMORI

Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan

`omori@e.u-tokyo.ac.jp`

ZHENGJUN ZHANG

Department of Statistics, University of Wisconsin Madison, WI 53706, USA

`zjz@stat.wisc.edu`

August, 2010

Abstract

Extreme values are often correlated over time, for example, in a financial time series, and these values carry various risks. Max-stable processes such as maxima of moving maxima (M3) processes have been recently considered in the literature to describe time-dependent dynamics, which have been difficult to estimate. This paper first proposes a feasible and efficient Bayesian estimation method for nonlinear and non-Gaussian state space models based on these processes and describes a Markov chain Monte Carlo algorithm where the sampling efficiency is improved by the normal mixture sampler. Furthermore, a unique particle filter that adapts to extreme observations is proposed and shown to be highly accurate in comparison with other well-known filters. Our proposed algorithms were applied to daily minima of high-frequency stock return data, and a model comparison was conducted using marginal likelihoods to investigate the time-dependent dynamics in extreme stock returns for financial risk management.

Key words: Bayesian analysis, Extreme value theory, Markov chain Monte Carlo, Marginal likelihood, Maxima of moving maxima processes, Stock returns.

1 Introduction

Extreme events rarely happen, but they could result in a major crisis. For example, huge shocks in stock markets or disastrous typhoons cause serious problems in daily life worldwide. It is critical to investigate the stochastics of rare events in detail, and therefore, increasing attention has been directed to risk analysis in various areas. The extreme value theory is known to be useful for describing the properties of extreme phenomena and is widely used in many fields ranging from environment science to insurance. For example, in financial econometrics, to measure market risk accurately, there are many concerns about estimating tails of the distributions of asset returns in the stock or exchange market. It is well known that for these returns, a normal distribution would fail to capture the extremal property because its tail is too thin. An extreme value distribution is a good alternative to investigate such risks. In the univariate case, the extreme value theory discusses such a distribution in terms of the limiting distribution of the maxima or the minima of a sequence of random variables. In the multivariate case, a component-wise maximum is used to study the statistical properties of its limiting distribution (e.g., Beirlant et al. (2004), Coles (2001), de Haan and Ferreira (2006), Embrechts et al. (1997), Resnick (1987, 2007) and Smith (2003)).

Financial time series data often exhibit clustering dynamics over time. Volatility clustering in stock returns is such an example, and the time-varying variance models such as the generalized autoregressive conditional heteroskedasticity (GARCH) and stochastic volatility models have been discussed extensively in the literature. On the other hand, the traditional extreme value theory focuses on independent random variables, but it is not sufficiently developed to describe the time-dependent structure of rare events despite its importance. For example, if the price of a stock drops drastically, a high degree of fluctuation in the price would persist over a few consecutive days. Without taking the time dependency of extreme values into account, we would fail to evaluate the financial risk.

To incorporate the dynamic structure of extremes, several extreme value models have been proposed to evaluate time-dependence in recent decades. One way to model such time-dependence is to exploit the state space representation in which parameters of popular extreme value models are assumed to follow autoregressive (AR) processes. An earlier

example was reported by Smith and Miller (1986), and many applications have been studied in various fields (e.g., Gaetan and Grigoletto (2004), Huerta and Sansó (2007)). Nakajima et al. (2009) used a different approach to express the generalized extreme value (GEV) random variable in a manner similar to that of a Box-Cox transformation where the state variables either follows an AR or a moving average (MA) process with Gumbel innovations.

Another way to capture time dependence is to consider max-stable processes, which are an infinite-dimensional generalization of the extreme value theory. Max-stable processes describe dependence across time, and several parametric models for max-stable processes have been proposed (e.g., Smith (2003)). The moving maxima (MM) process is a stationary stochastic sequence that marginally follows the Fréchet distribution. Davis and Resnick (1989) proposed the max-autoregressive moving average (MARMA) process, and Davis and Resnick (1993) discussed its prediction in detail. Deheuvels (1983) defined the class of maxima of moving maxima (M3) processes, and Smith and Weissman (1996) extended the M3 process to the multivariate maxima of moving maxima (M4) process. Smith and Weissman (1996) also showed that a large class of max-stable processes with unit Fréchet margins could be approximated by M4 processes under quite general conditions. Zhang and Smith (2010) also discussed this approximation in detail. Although this approach is promising, it is difficult to estimate parameters, thus there have not yet been many applications of these models to real data. Estimation methods and many properties of these processes have been explored, for example, by Chamú Morales (2005), Hall et al. (2002), Heffernan et al. (2007), Martins and Ferreira (2005), Smith (2003), Zhang (2002), Zhang (2009), Zhang and Smith (2004) and Zhang and Smith (2010).

Max-stable processes can be expressed as a nonlinear state space model, and a simple Markov chain Monte Carlo (MCMC) estimation method for such a model was proposed in Chamú Morales (2005). However, an efficient estimation method has not yet been proposed. This statistical model has an identifiability problem, which is well known as a label switching problem in finite mixture and Markov switching models (e.g., Frühwirth-Schnatter (2006)). This problem may result in the slow mixing property of the Markov chains and misleading conclusions. To overcome these difficulties, we propose a feasible and efficient MCMC algorithm where we sample state variables by using a mixture of normal distributions. Fur-

thermore, based on the mixture sampling method, we propose a novel and efficient particle filter method that adapts to extreme changes in the dependent variables. Furthermore, we demonstrate that popular filters such as auxiliary particle filters (e.g., Pitt and Shephard (1999)) fail to capture them and that they give misleading inaccurate estimates in our empirical examples. Concerning label switching, this paper solves the identifiability problem by applying the permutation sampler proposed by Früwirth-Schnatter (2001). The proposed methods are applied to high frequency stock data and compared to a time-dependent GEV model based on max-stable processes with a simple GEV model using marginal likelihoods.

The rest of paper is organized as follows. Section 2 reviews statistical models based on max-stable processes. Section 3 describes our feasible efficient MCMC algorithm for nonlinear and non-Gaussian state space models where the state variables follow the M3 processes. A particle filter that adapts to extreme observations is also proposed to compute the likelihood function. Section 4 illustrates the high efficiency of our sampling algorithm using simulated data. In Section 5, we apply our proposed method to daily minima of intraday stock returns using high frequency stock price data and compare models using the marginal likelihoods. Further, the high performance of our proposed particle filter is shown in comparison with several other well-known particle filters. Finally, Section 6 concludes this paper.

2   Modeling of max-stable processes

We first reviewed max-stable processes that form the basis of statistical models for extreme values with dependent structure (e.g., Smith (2003)). We assumed the marginal distributions are unit Fréchet, since other random variables that follow extreme value distribution can be transformed to follow unit Fréchet. A stochastic process $\{X_t,\ t = 1, 2, ...\}$ with unit Fréchet margins is called max-stable if any finite dimensional distributions are max-stable, $i.e.$,

$$\Pr\{X_1 \leq nx_1, ..., X_r \leq nx_r\}^n = \Pr\{X_1 \leq x_1, ..., X_r \leq x_r\} \ \text{ for all } n \geq 1, r \geq 1.$$

This definition can be extended to multivariate D-dimensional processes. A stochastic process $\{X_{t,d},\ t = 1, 2, ...,\ 1 \leq d \leq D\}$ with unit Fréchet margins is max-stable if for any

$n \geq 1,\ r \geq 1,$

$$\Pr\{X_{t,d} \leq nx_{t,d},\ 1 \leq t \leq r,\ 1 \leq d \leq D\}^n = \Pr\{X_{t,d} \leq x_{t,d},\ 1 \leq t \leq r,\ 1 \leq d \leq D\}.$$

Next, we introduce two parametric models based on these max-stable processes:

(i) Moving maximum (MM) process. It is given by

$$X_t = \max_{-\infty < k < \infty} \alpha_k Z_{t-k},$$

where $\{\alpha_k\}$ is a nonnegative sequence where $\sum_k \alpha_k = 1$ and Z_t's are i.i.d. unit Fréchet random variables. We can show that X_t is a stationary stochastic sequence and the marginal distributions are unit Fréchet.

(ii) Maxima of moving maxima (M3) process. Deheuvels (1983) defined the process as

$$X_t = \max_{0 \leq l}\ \max_{-\infty < k < \infty} \alpha_{l,k} Z_{l,t-k},$$

where $\{\alpha_{l,k}\}$ is a nonnegative sequence satisfying $\sum_l \sum_k \alpha_{l,k} = 1$ and $Z_{l,t}$'s are i.i.d. unit Fréchet random variables. The M3 process is also stationary and max-stable. The key characterization of M3 process in practice is that any stationary max-stable process with unit Fréchet margins can be approximated arbitrarily closely by the M3 process under general conditions (see Zhang and Smith (2010) for details).

In practice, we set l and k to some finite constants when we estimate parameters of these processes.

3 The GEV-M3 model

3.1 Model specification

We set X as a random variable with unit Fréchet distribution function given by

$$P(X \leq x) = \begin{cases} \exp(-x^{-1}), & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{1}$$

and defined Y as a function of X such that

$$Y = \mu + \psi \frac{X^\xi - 1}{\xi}, \quad \mu \in R,\ \psi > 0,\ \xi \in R. \tag{2}$$

Then, Y follows GEV distribution with a distribution function

$$P(Y \leq y) = \exp \left\{ - \left(1 + \xi \frac{y - \mu}{\psi} \right)_+^{-1/\xi} \right\}, \tag{3}$$

where $y_+ = \max(y, 0)$. We considered the state space model where the measurement equation is (2) with an idiosyncratic observational shock. The state equation was the M3 process and the state variable marginally follows the unit Fréchet distribution (e.g., Chamú Morales (2005)):

GEV-M3 model:

$$y_t = \mu + \psi \frac{x_t^\xi - 1}{\xi} + \epsilon_t, \quad t = 1, 2, ..., n, \tag{4}$$

$$x_t = \max_{0 \leq l \leq L} \max_{0 \leq k \leq K} \alpha_{l,k} z_{l,t-k}, \quad t = 1, 2, ..., n, \tag{5}$$

where $\{z_{l,t}\}$ is a sequence of i.i.d. unit Fréchet random variables, $\{\epsilon_t\}$ is a sequence of i.i.d. normally distributed random variables with mean 0, variance σ^2 and $\{\alpha_{l,k}\}$ is a sequence of nonnegative constants satisfying $\sum_l \sum_k \alpha_{l,k} = 1$. It is more realistic to expect the maxima or minima of a large set of data to have GEV marginal distribution instead of unit Fréchet distribution. In addition, the independent GEV distribution model is nested in this representation.

As Hall et al. (2002) mentioned, we attached more importance to current and past disturbances than to future disturbances and hence set $\alpha_{l,k} = 0$ for $k < 0$. We refer to this model as the GEV-M3 model. Note that $y_t \sim \mathrm{GEV}(\mu, \psi, \xi)$ when there is no observational noise.

3.2   Bayesian estimation

Taking a Bayesian approach, we proposed a feasible efficient MCMC algorithm for estimating parameters and state variables in the GEV-M3 model (4) and (5). The unknown model parameters of the GEV-M3 model are equal to $\vartheta \equiv (\lambda, \sigma^2, \alpha)$, where $\lambda = (\mu, \psi, \xi)'$. Regarding prior distributions, we assumed prior independence among λ, σ^2 and α, *i.e*, $\pi(\lambda, \sigma^2, \alpha) = \pi(\lambda)\pi(\sigma^2)\pi(\alpha)$. Concerning σ^2, we used the conditionally conjugate prior $\sigma^2 \sim \mathrm{IG}(n_0/2, S_0/2)$, where IG denotes the inverse gamma distribution and, n_0 and S_0 are hyperparameters. Since no such conditionally conjugate priors can be found for λ and α, the choice of $\pi(\lambda)$ and $\pi(\alpha)$ was completely arbitrary. In our studies, we assumed prior independence among all components of λ, with μ and ξ following a normal and ψ following a Gamma distribution.

For practical Bayesian estimation, we used MCMC methods to sample from the posterior distribution, (e.g., Chib (2001), Chib and Greenberg (1996), Gamerman and Lopes (2006), Geweke (2005) and Koop (2003)). Concerning state variables, since Markov properties of general M3 processes are not well known, which makes it difficult to sample $x \equiv \{x_t\}$ directly, we therefore considered sampling $z \equiv \{z_{l,t}\}$. As is common for state space models, the data augmentation method was used by introducing the latent state process z as missing data.

Firstly, we set $y = \{y_t\}$ and considered the following MCMC algorithm to sample from the joint posterior density $\pi(\vartheta, z|y)$.

Algorithm 1: MCMC algorithm for the GEV-M3 model

1. Generate $(\mu, \psi, \xi)|\sigma^2, x, y$.

2. Generate $\sigma^2|\mu, \psi, \xi, x, y$.

3. Generate $\alpha|\mu, \psi, \xi, \sigma^2, z, y$.

4. Generate $z|\vartheta, y$.

It is easy to sample $\sigma^2 \,|\, \mu, \psi, \xi, x, y \sim \text{IG}(\hat{n}/2, \hat{S}/2)$ in Step 2 where $\hat{n} = n_0 + n$, $\hat{S} = S_0 + \sum_{t=1}^{n} \{y_t - \mu - \psi(x_t^\xi - 1)/\xi\}^2$, while sampling from $(\mu, \psi, \xi) \,|\, \sigma^2, x, y$ and $\alpha \,|\, \mu, \psi, \xi, \sigma^2, z, y$ in Steps 1 and 3 requires the implementation of a Metropolis-Hastings (MH) algorithm. To sample from $(\mu, \psi, \xi) \,|\, \sigma^2, x, y$, we generated a candidate from a normal distribution, where we set its mean and covariance matrix to be equal to the mode and the inverse of the Hessian matrix multiplied by -1 of the conditional posterior densities (see Appendix A.1). Concerning $\alpha \,|\, \mu, \psi, \xi, \sigma^2, z, y$, we implemented a random walk MH algorithm where a candidate was generated from a normal distribution truncated over the region $R = \{\alpha : \alpha_{l,k} \geq 0, \sum_l \sum_k \alpha_{l,k} = 1\}$.

Since the GEV-M3 model is a non-linear and non-Gaussian state space model, it is difficult to sample the state variables. Moreover, since the likelihood is invariant to relabeling the component l in the M3 process, it is likely that the labeling of the unobserved indices changes in the course of sampling from the posterior distribution. Label switching, which is a jump between the various labeling subspaces, is well known in the finite mixture and Markov switching models (Frühwirth-Schnatter (2006)) and can make the mixing of Markov chain very slow, leading to incorrect conclusions.

Considering these factors, our paper employed an auxiliary mixture sampling and approximated the non-linear non-Gaussian state space models (4) and (5) by a very accurate finite mixture of non-linear Gaussian state space models. Thanks to this fine approximation, we could sample the state variables from their posterior distribution efficiently through the

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).

MCMC algorithm. We will discuss this in more detail in the next subsection. Concerning the identification, we used the random permutation sampler by Früwirth-Schnatter (2001), which leads to balanced label switching by concluding each MCMC draw by a randomly selected permutation of the labeling. This method can be used to obtain a sample that jumps between the various subspaces in a balanced fashion. The MCMC output of the random permutation sampler was explored to find a suitable identifiability constraint.

3.2.1 Auxiliary mixture sampler for efficient sampling of state variables

As an efficient sampling method, auxiliary mixture sampling has been well developed in the context of the stochastic volatility model by approximating the log χ_1^2 density by a finite normal mixture (Kim et al. (1998) and Omori et al. (2007)). Recently, this idea was extended to efficient estimation of non-Gaussian models with latent variables 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 et al. (2009) and Frühwirth-Schnatter and Wagner (2006)). Moreover, Nakajima et al. (2009) applied this approximation method to extreme value analysis.

Firstly, we transform unit Fréchet random variable $z_{l,t}$ in the state equation (5) to $w_{l,t} \equiv \log z_{l,t}$. Then $w_{l,t}$ follows the Gumbel distribution given by

$$G(x) \equiv \Pr(w_{l,t} \leq x) = \exp(-\exp(-x)),$$

and we replaced $z_{l,t}$ in the equation (5) by $\exp(w_{l,t})$, that is,

$$x_t = \max_{0 \leq l \leq L} \max_{0 \leq k \leq K} \alpha_{l,k} \exp(w_{l,t-k}). \tag{6}$$

We approximated the exact probability density function $g(w_{l,t})$ of the Gumbel distribution by a normal mixture of S components as

$$g(w_{l,t}) = \exp(-w_{l,t} - e^{-w_{l,t}}) \approx \hat{g}(w_{l,t}) = \sum_{i=1}^{S} p_i f_N(w_{l,t}|m_i, v_i^2), \tag{7}$$

where $f_N(w_{l,t}|m_i, v_i^2)$ denotes a probability density function of a normal distribution with mean m_i and variance v_i^2. Frühwirth-Schnatter and Frühwirth (2007) proposed the fine mixture approximation based on $S = 10$ components where the selection of (p_i, m_i, v_i^2) for $i = 1, \ldots, 10$ is reproduced in Table 1.

Next, we introduced a mixture indicator variable, $s_{l,t} \in \{1, \ldots, S\}$ for $t = 0, \ldots, n$, $l = 0, \ldots, L$. Conditional on $s \equiv \{s_{l,t}\}$, equations (4) and (5) form a non-linear Gaussian state space model,

$$y_t = \mu + \psi\frac{x_t^\xi - 1}{\xi} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad t = 1, 2, ..., n, \tag{8}$$

$$x_t = \max_{0 \le l \le L} \max_{0 \le k \le K} \alpha_{l,k} \exp(w_{l,t-k}), \tag{9}$$

$$w_{l,t} = m_{s_{l,t}} + v_{s_{l,t}} u_{l,t}, \quad u_{l,t} \sim N(0, 1), \quad t = 0, ..., n, \quad l = 0, ..., L. \tag{10}$$

Taking the fine approximation into consideration, we set $w = \{w_{l,t}\}$ and implemented the improved MCMC algorithm to sample from the joint posterior density $\pi(\vartheta, s, w|y)$.

Algorithm 2: Improved MCMC algorithm for the GEV-M3 model

1. Generate $(\mu, \psi, \xi)|\sigma^2, x, y$.

2. Generate $\sigma^2|\mu, \psi, \xi, x, y$.

3. Generate $\alpha|\mu, \psi, \xi, \sigma^2, w, y$.

4. Generate $w|\vartheta, s, y$.

5. Generate $s|w$

6. Permute index l randomly by

   (a) Selecting randomly one of the $(L + 1)!$ possible permutations $\rho(0), ..., \rho(L)$,

   (b) Replacing $\alpha_{l,k}$, $w_{l,t}$, $s_{l,t}$ by $\alpha_{\rho(l),k}$, $w_{\rho(l),t}$, $s_{\rho(l),t}$ respectively for $k = 0, ..., K$, $t = 0, ..., n$.

Sampling the latent mixture indicator variables s is a standard step in finite mixture modeling (see e.g., Frühwirth-Schnatter (2006)). Regarding sampling the latent state process w, we firstly divided the domain of the state variable $w_{l,t}$ into finite intervals depending on its posterior representation. In each interval, we approximated the posterior density by normal density based on a Taylor expansion of the log density in the observation equation (4). Then, we combined all proposals into a mixture normal density. Details are provided in Appendix A.1.

3.2.2 Reweighting to correct for the mixture approximation error

Although the normal mixture distribution approximates the Gumbel distribution well, the MCMC samples drawn from the approximate distribution $\hat{\pi}(\vartheta, w|y)$ by Algorithm 2 is different from those from the exact posterior distribution $\pi(\vartheta, w|y)$. Here, $\hat{\pi}(\vartheta, w|y)$ is the marginal posterior of the approximate model, where the exact density $g(w_{l,t})$ is substituted by the approximate density $\hat{g}(w_{l,t})$ given by (7).

This subsection describes the method of correcting for the minor approximation error. ϑ^j and w^j denoted the j-th sample from the approximated model, $j = 1, \ldots, M$, where M is the number of iterations. To sample from the exact posterior distribution $\pi(\vartheta, w|y)$, we resampled the draws from the approximate posterior density with weights proportional to

$$\eta_j = \frac{\eta_j^*}{\sum_{i=1}^M \eta_i^*}, \quad \eta_j^* = \frac{\pi(\vartheta^j, w^j|y)}{\hat{\pi}(\vartheta^j, w^j|y)} = \frac{f(w^j)}{\hat{f}(w^j)}, \quad j = 1, \ldots, M, \tag{11}$$

where $f(w) = \prod_{t=0}^n \prod_{l=0}^L g(w_{l,t})$ is given by the product of the exact densities and $\hat{f}(w) = \prod_{t=0}^n \prod_{l=0}^L \hat{g}(w_{l,t})$ is given by the product of the approximate densities, i.e. $\hat{g}(w_{l,t}) = \sum_{s_{l,t}=1}^S p_{s_{l,t}} f_N(w_{l,t}|m_{s_{l,t}}, v_{s_{l,t}}^2)$. The posterior moments were obtained by computing the weighted average of the MCMC draws (e.g., Kim et al. (1998)).

3.3 Associated efficient particle filter

In this subsection, an efficient particle filter method is proposed to compute the likelihood, which is required to conduct a model comparison discussed in Subsection 5.4. Here, we assumed ϑ to be known. The basic idea was to sample from a target posterior distribution recursively with the help of the importance function that approximates the target density well. For the GEV-M3 model, the measurement density and the evolution density were respectively $f(y_t|z_{t-K:t}, \vartheta) = f(y_t|x_t, \vartheta)$ from (4) and $f(z_t) \equiv \prod_{l=0}^L f(z_{l,t})$ from (5) where $z_t = (z_{0,t}, \ldots, z_{L,t})$, $z_{t-K:t} = \{z_j\}_{j=t-K}^t$, $x_t \equiv \max_{0 \le l \le L} \max_{0 \le k \le K} \alpha_{l,k} z_{l,t-k}$. The

associated particle filter is based on

$$f(z_{t+1}, z_{t+1-K:t}|y_{1:t+1}, \vartheta) \propto f(y_{t+1}|x_{t+1}, \vartheta)f(z_{t+1})f(z_{t+1-K:t}|y_{1:t}, \vartheta),$$

where $y_{1:t} = \{y_j\}_{j=1}^t$. Here, we assumed $K \geq 1$ because it is not necessary to consider the sequential filtering in the case of $K = 0$. To draw particles from $f(z_{t+1-K:t}|y_{1:t}, \vartheta)$, we assumed it was approximated well by $\hat{f}(z_{t+1-K:t}|y_{1:t}, \vartheta)$ based on discrete distribution functions.

The simple particle filter (PF) uses $f(z_{t+1})$ as an importance function, but it is known to produce an inaccurate estimate of the likelihood. Alternatively, the auxiliary particle filter (APF, Pitt and Shephard (1999)) is well known as an efficient filter in various fields. However, such a filter often generates particles with almost zero importance weights for extreme observations. This is because the APF generates particles without considering that extreme values are observed occasionally. Many particles with zero weights result in the poor approximation of the filtering density and inaccurate estimation of the likelihood.

To solve these problems, we proposed an efficient particle filter that always generates particles directly using the information y_{t+1}. In the first step, we generated $z_{t+1-K:t}^i$ from $\hat{f}(z_{t+1-K:t}|y_{1:t}, \vartheta)$. In the second step, we assumed x_{t+1} is driven by the l^*-th element among $L + 1$ sequences that comprise the M3 process. Then, the sequence with l^* was supposed to link deeply with the observation y_{t+1}. Since the situation where $l \in \{0, \ldots, L\}$ corresponds to l^* was not observable, we sampled l^* with each weight $\alpha_l^+ = \sum_{k=0}^K \alpha_{l,k}$ (see Chamú Morales (2005)). This weight can be interpreted as the proportion of times that the process x_{t+1} is driven by the lth signature pattern that specifies the shape of the process near its local maximum (see e.g., Smith (2003) and Smith and Weissman (1996)). In the third step, $z_{l,t+1}$, $l \neq l^*$ was generated from the unit Fréchet distribution. Finally, given $z_{t+1-K:t}^i$, l^* and $z_{l,t+1}$, we sample $z_{l^*,t+1}$ based on the mixture approximation of Gumbel density (see Appendix A.2). Therefore, our proposed importance function is given by

$$
\begin{aligned}
g(z_{t+1}, z_{t+1-K:t}^i|y_{1:t+1}, \vartheta) &= g(z_{t+1}|y_{t+1}, z_{t+1-K:t}^i, \vartheta)\hat{f}(z_{t+1-K:t}^i|y_{1:t}, \vartheta) \\
&= \sum_s g(z_{t+1}|y_{t+1}, z_{t+1-K:t}^i, \vartheta, s)\pi(s)\hat{f}(z_{t+1-K:t}^i|y_{1:t}, \vartheta)
\end{aligned}
$$

where

$$
\begin{aligned}
&g(z_{t+1}|y_{t+1}, z_{t+1-K:t}^i, \vartheta, s) \\
&= \sum_{l^*=0}^L \alpha_{l^*}^+ g(z_{l^*,t+1}|y_{t+1}, z_{l,t+1}, z_{t+1-K:t}, s_{l^*,t+1}, \vartheta) \prod_{l \neq l^*} f(z_{l,t+1}), \qquad (12)
\end{aligned}
$$

$\pi(s) = \prod_l \pi(s_{l,t+1})$ and $g(z_{l^*,t+1}|y_{t+1}, z_{l,t+1}, z_{t+1-K:t}, s_{l^*,t+1}, \vartheta)$ comprise the a mixture normal distribution discussed in Appendix A.2. To summarize, we implemented the following algorithm to generate $(z_{t+1}, z_{t+1-K:t})$ from our proposed importance function:

Algorithm 3: Generating particles from our importance function

1. Generate $z_{t+1-K:t}^i$ from $\hat{f}(z_{t+1-K:t}|y_{1:t}, \omega)$.

2. Generate l^* from the set $\{0, \ldots, L\}$ with weight $\alpha_l^+ = \sum_{k=0}^{K} \alpha_{l,k}$.

3. Generate $z_{l,t+1}$, $l \neq l^*$ from the unit Fréchet distribution.

4. Generate $z_{l^*,t+1}$ from $g(z_{l^*,t+1}|y_{t+1}, z_{l,t+1}, z_{t+1-K:t}, s_{l^*,t+1}, \vartheta)$.

If y_{t+1} is large, more particles are generated from the tail of unit Fréchet distribution, while a moderate y_{t+1} value corresponds to the generation of particles from around the mode of unit Fréchet. Namely, our method automatically adapts the importance function to the value of y_{t+1}.

Taking that into consideration, we propose the following particle filter method:

1. Initialize $t = 1$, generate z_0^i and z_1^i from unit Fréchet for $i = 1, \ldots, I$.

   (a) Compute $w_1^i = f(y_1|x_1^i, \vartheta)$ and $W_1^i = F(y_1|x_1^i, \vartheta)$, where F denotes the distribution function of y_t given x_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 $\hat{f}(z_1^i|y_1, \vartheta) = w_1^i / \sum_{j=1}^{I} w_1^j$, $i = 1, \ldots, I$.

2. Generate $(z_{t+1}^i, z_{t+1-K:t}^i)$ from the importance function $g(z_{t+1}, z_{t+1-K:t}|y_{1:t+1}, \vartheta)$ for $i = 1, \ldots, I$.

   (a) Compute

   $$w_{t+1}^i = \frac{f(y_{t+1}|x_{t+1}^i, \vartheta)f(z_{t+1}^i)\hat{f}(z_{t+1-K:t}^i|y_{1:t}, \vartheta)}{g(z_{t+1}^i, z_{t+1-K:t}^i|y_{1:t+1}, \vartheta)} = \frac{f(y_{t+1}|x_{t+1}^i, \vartheta)\hat{f}(z_{t+1}^i)}{g(z_{t+1}^i|y_{t+1}, z_{t+1-K:t}^i, \vartheta)},$$

   $$W_{t+1}^i = \frac{F(y_{t+1}|x_{t+1}^i, \vartheta)f(z_{t+1}^i)}{g(z_{t+1}^i|y_{t+1}, z_{t+1-K:t}^i, \vartheta)}, \quad i = 1, \ldots, I,$$

   and save $\bar{w}_{t+1} = \frac{1}{I} \sum_{i=1}^{I} w_{t+1}^i$, $\bar{W}_{t+1} = \frac{1}{I} \sum_{i=1}^{I} W_{t+1}^i$.

   (b) Set $\hat{f}(z_{t+2-K:t+1}^i|y_{1:t+1}, \vartheta) = w_{t+1}^i / \sum_{j=1}^{I} w_{t+1}^j$, $i = 1, \ldots, I$.

3. Go to 2.

It can be shown that as $I \to \infty$, $\bar{w}_{t+1} \xrightarrow{p} f(y_{t+1}|y_{1:t}, \vartheta)$ and $\bar{W}_{t+1} \xrightarrow{p} F(y_{t+1}|y_{1:t}, \vartheta)$ (e.g., Doucet et al. (2001)). Then, it follows that $\sum_{t=1}^{n} \log \bar{w}_t \xrightarrow{p} \sum_{t=1}^{n} \log f(y_t|y_{1:t-1}, \vartheta)$. In Subsection 5.4, we show that our proposed filter outperforms other well-known filters.

4 Illustrative examples

4.1 Evidence of high efficiency in sampling $z_{l,t}$

This section shows the high efficiency of our proposed estimation method using simulation data. We assumed $l = 0, 1$ because we expected shocks in such as stock markets consist of transitory and persistent components in the M3 process, which corresponds to weak and strong time-dependent structures. We generated 1,000 observations from the following GEV-M3 model:

$$y_t = \mu + \psi \frac{x_t^{\xi} - 1}{\xi} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2) \quad t = 1, ..., n, \tag{13}$$

$$x_t = \max \begin{pmatrix} \alpha_{0,0} z_{0,t}, & \alpha_{0,1} z_{0,t-1} \\ \alpha_{1,0} z_{1,t}, & \alpha_{1,1} z_{1,t-1} \end{pmatrix}, \tag{14}$$

where we set true values as $\mu = 0.2$, $\psi = 0.1$, $\xi = 0.3$, $\sigma = 0.05$, $(\alpha_{0,0}, \alpha_{1,0}, \alpha_{0,1}, \alpha_{1,1}) = (0.6, 0.2, 0.01, 0.19)$. And Figure 1 shows simulation data of y.
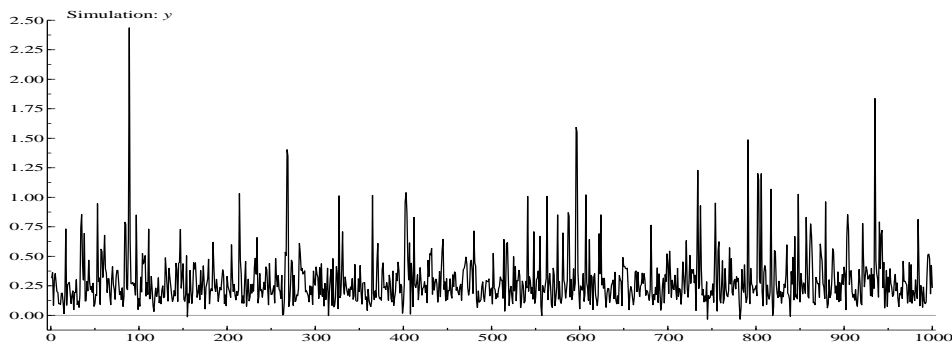

Figure 1: Simulation data of y.

Since huge state variables are directly linked with extreme observations and have a much larger impact on the process, it is important to estimate those generated from the tail of the unit Fréchet distribution. To investigate the sampling efficiency in such a case, we focused on large state variables. We fixed all parameters except $z_{l,t}$ in GEV-M3 model (13), (14) and sampled 10,000 MCMC draws by our proposed method after the initial 1,000 samples were discarded as a burn-in period. In our method, MCMC draws were resampled to correct approximation errors discussed in Subsection 3.2.2[1]. As a benchmark, we compared our method with the simple algorithm where unit Fréchet distribution was used as a proposal distribution. We drew 1,000,000 samples using the simple algorithm after the initial 100,000 samples were discarded as a burn-in period.

---

[1]Inefficiency factors and acceptance rates are computed using original samples.

Tables 2 (the benchmark sampler) and 3 (our sampler) give the estimates of posterior means, standard deviations, the 95% credible intervals, inefficiency factors and acceptance rates for the maximum of state variables $z_{0,89}$, the second largest $z_{1,596}$, the third largest $z_{1,268}$, and around the 99th percentile $z_{1,942}$ and the 95th percentile $z_{0,525}$.

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 (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 the inefficiency factor is also known as relative numerical efficiency (Geweke (1992)). When the inefficiency factor is equal to m, we need to draw the MCMC sample m times more than the uncorrelated sample [2].

Parameter	True	Mean	Stdev.	95% interval	Inefficiency	Acceptance rate
$z_{0,89}$	1486.8	1510.5	112.9	[1320.1, 1758.2]	8862.9	0.0001
$z_{1,596}$	1161.1	1182.4	104.7	[1029.4, 1376.3]	4869.2	0.002
$z_{1,268}$	811.3	780.7	61.46	[665.3, 907.4]	4772.3	0.0003
$z_{1,942}$	106.1	101.3	29.33	[1.067, 138.2]	332.8	0.066
$z_{0,525}$	19.65	13.10	3.584	[7.043, 21.00]	29.1	0.062

Table 2: Estimation result (the benchmark sampler).

Parameter	True	Mean	Stdev.	95% interval	Inefficiency	Acceptance rate
$z_{0,89}$	1486.8	1508.4	92.51	[1334.1, 1689.7]	2.2	0.625
$z_{1,596}$	1161.1	1178.3	80.28	[1028.2, 1337.7]	2.2	0.627
$z_{1,268}$	811.3	783.7	61.27	[670.9, 908.3]	2.2	0.637
$z_{1,942}$	106.1	108.1	16.26	[78.74, 140.2]	4.6	0.608
$z_{0,525}$	19.65	13.13	3.563	[7.304, 20.93]	3.1	0.609

Table 3: Estimation result (our sampler).

Although all true values are contained in the 95% credible intervals for both methods, inefficiency factors in our method (2.2 ∼ 4.6) were much smaller than those in the benchmark method (29.1 ∼ 8862.9). Acceptance rates in the benchmark method were very small, and Figure 2 shows that candidates are rejected too often and Markov chains do not mix well. On the other hand, in our method, acceptance rates were about 0.6, and it is clear from Figure 3 that the sample paths are stable. In addition, the sample autocorrelations decayed very slowly in the benchmark sampler (Figure 4), while they vanished quickly in our sampler (Figure 5), which indicates that our sampling method efficiently produces uncorrelated samples. This is because our method generates samples that considered the data information. On the other hand, the benchmark does not use this information, even though large state variables are highly related to extreme observations.

---

[2] The inefficiency factors are computed using bandwidths $b_w = 20,000$ and 150 for the benchmark sampler and our sampler, respectively.
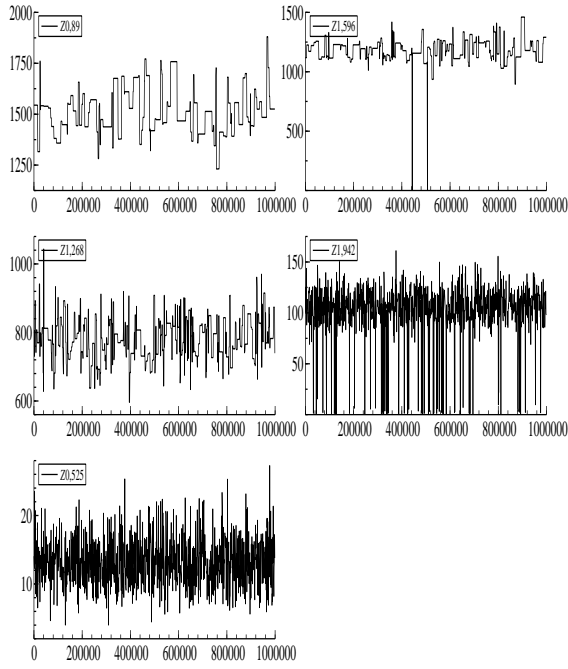
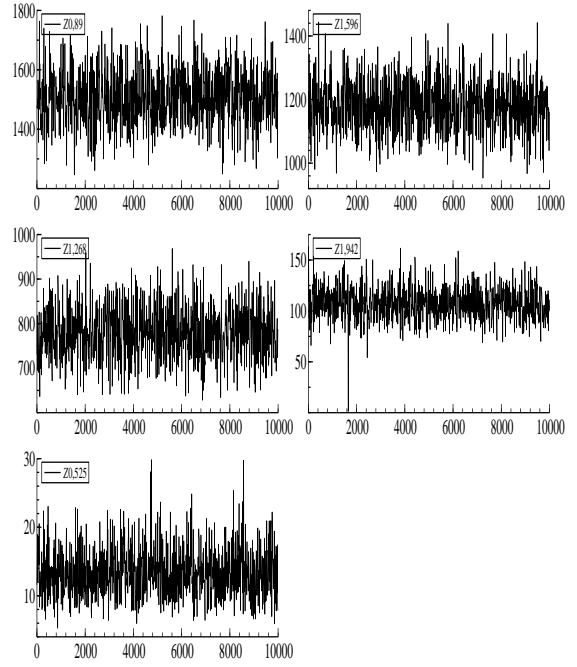Figure 2: Sample paths (the benchmark sampler).

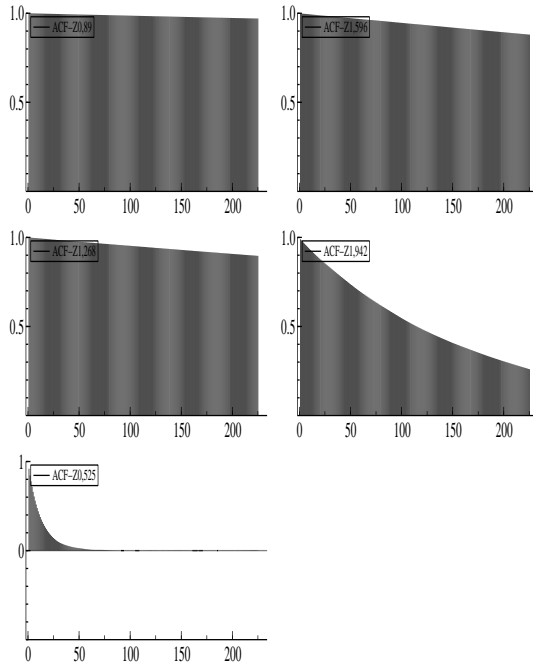Figure 3: Sample paths (our sampler).

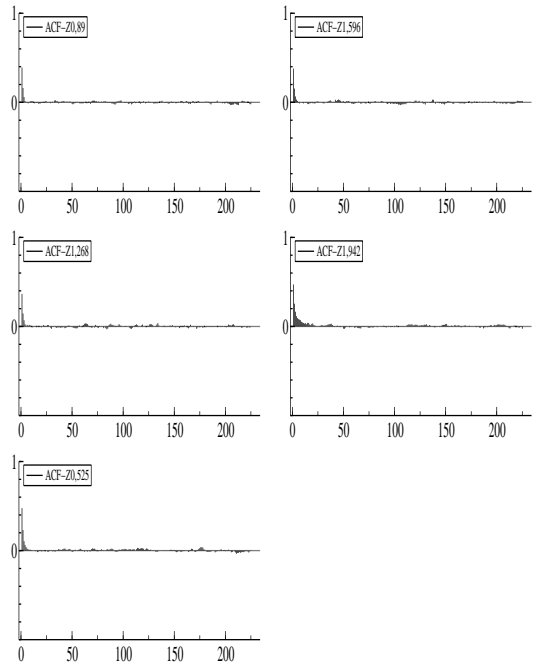Figure 4: Sample autocorrelations (the benchmark sampler).

Figure 5: Sample autocorrelations (our sampler).

4.2 Overall sampling efficiency

Next we considered estimating not only state variables but also all parameters. It will be shown that our proposed method works well and outperforms the benchmark sampler. Again, we used the data set generated from the GEV-M3 models (13) and (14) [3]. Here, we assumed the following prior distributions:

$$\mu \sim \mathrm{N}(0, 10), \quad \psi \sim \mathrm{G}(2, 2), \quad \xi \sim \mathrm{N}(0, 4)$$

$$\sigma^{-2} \sim \mathrm{Ga}(2.5, 0.025), \quad (\alpha_{0,0}, \alpha_{1,0}, \alpha_{0,1}, \alpha_{1,1}) \sim \mathrm{Dirichlet}(1, 1, 1, 1).$$

We drew $M = 200,000$ samples using our proposed method after the initial 20,000 samples were discarded as a burn-in period. In the first step, we drew samples using a random permutation sampler. Figure 6 shows the scatter plot of α, which indicates $\alpha_{0,0} > \alpha_{1,0}$ is a suitable constraint for identifiability of the index l. To compare the sampling efficiencies, we again adopted the algorithm as a benchmark where unit Fréchet distribution was simply used as a proposal distribution for all state variables. We sampled $M = 4,000,000$ draws by the benchmark method after the initial 400,000 samples were discarded as a burn-in period.
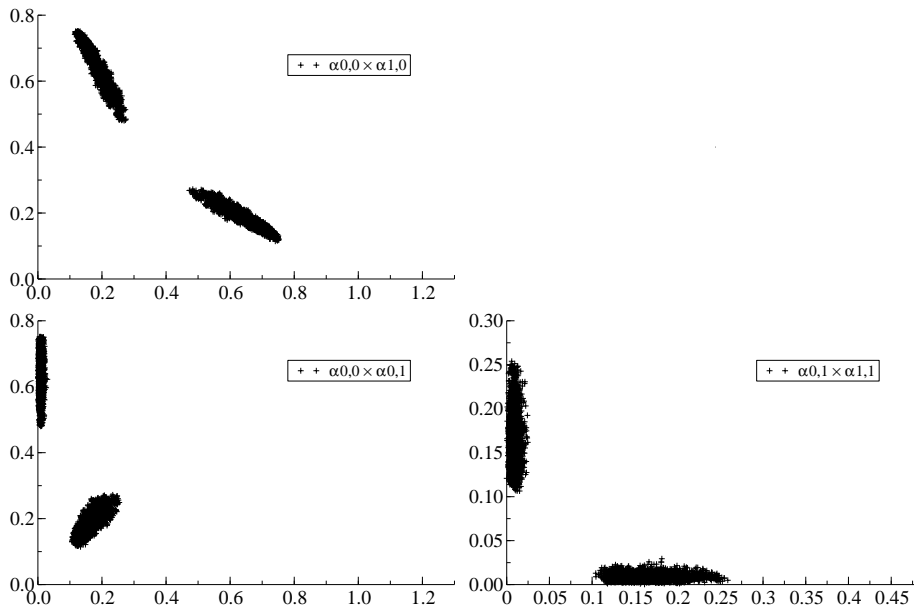

Figure 6: Scatterplot of MCMC draws α from random permutation sampler for simulation data (x-label $\times$ y-label).

---

[3] The data set is different from that in Subsection 4.1 because some state variables in the previous data set are too large for the benchmark method to converge the Markov chain.

The sample paths and the sample autocorrelation function are given in Figures 7 and 9 for the benchmark method and Figures 8 and 10 for our method. Tables 4 (the benchmark sampler) and 5 (our sampler) show the estimates for posterior means, standard deviations, the 95% credible intervals and inefficiency factors [4]. Here, $z_{\max}$ denotes the maximum of $\{z_{l,t}\}$. In the benchmark, all true values were contained in the 95% credible intervals as well as in our proposed method, but the inefficiency factors were very large and the sample autocorrelations decayed very slowly. Therefore, the poor estimation for state variables crucially affects the efficiency of sampling other parameters. On the other hand, these results show that the inefficiency factors in our sampler were much smaller and that our autocorrelations vanished more quickly, which implies our method outperforms the benchmark sampler.

Parameter	True	Mean	Stdev.	95% interval	Inefficiency
μ	0.2	0.196	0.004	[0.188, 0.205]	1114.6
ψ	0.1	0.090	0.006	[0.077, 0.102]	6177.2
ξ	0.3	0.289	0.049	[0.200, 0.399]	75218.
σ	0.05	0.058	0.005	[0.048, 0.069]	6686.7
$\alpha_{0,0}$	0.6	0.649	0.058	[0.532, 0.762]	14658.
$\alpha_{1,0}$	0.2	0.182	0.033	[0.119, 0.249]	14355.
$\alpha_{0,1}$	0.01	0.008	0.003	[0.003, 0.017]	68135.
$\alpha_{1,1}$	0.19	0.159	0.028	[0.105, 0.216]	13098.
$z_{\max}$	497.8	809.7	412.8	[304.3, 1916.8]	84639.

Table 4: Estimation result for simulated data (the benchmark sampler).

Parameter	True	Mean	Stdev.	95% interval	Inefficiency
μ	0.2	0.196	0.004	[0.188, 0.204]	51.0
ψ	0.1	0.090	0.006	[0.077, 0.102]	172.6
ξ	0.3	0.284	0.047	[0.197, 0.383]	345.5
σ	0.05	0.057	0.005	[0.047, 0.069]	292.7
$\alpha_{0,0}$	0.6	0.669	0.056	[0.556, 0.776]	1305.7
$\alpha_{1,0}$	0.2	0.171	0.032	[0.111, 0.237]	1256.1
$\alpha_{0,1}$	0.01	0.008	0.003	[0.003, 0.016]	294.0
$\alpha_{1,1}$	0.19	0.150	0.026	[0.099, 0.204]	1172.2
$z_{\max}$	497.8	794.6	424.4	[324.7, 1864.3]	314.1

Table 5: Estimation result for simulated data (our sampler).

---

[4]The inefficiency factors are computed using bandwidths $b_w = 250,000$ and 3500 for the benchmark sampler and our sampler.
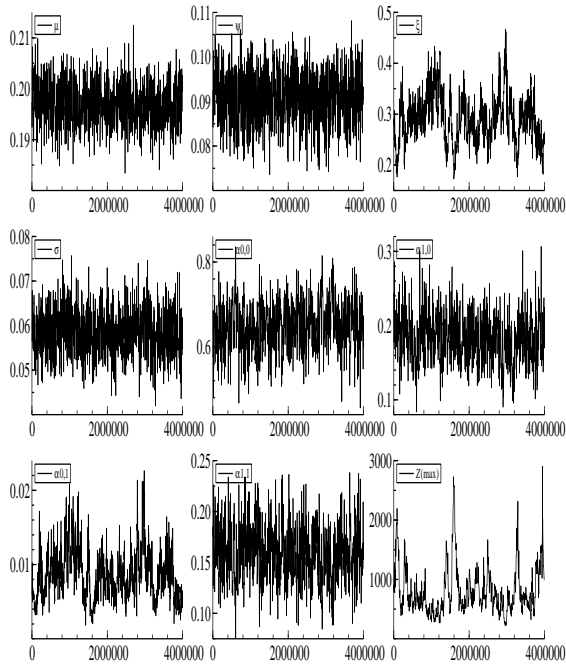
Figure 7: Sample paths for simulation data (the benchmark sampler).
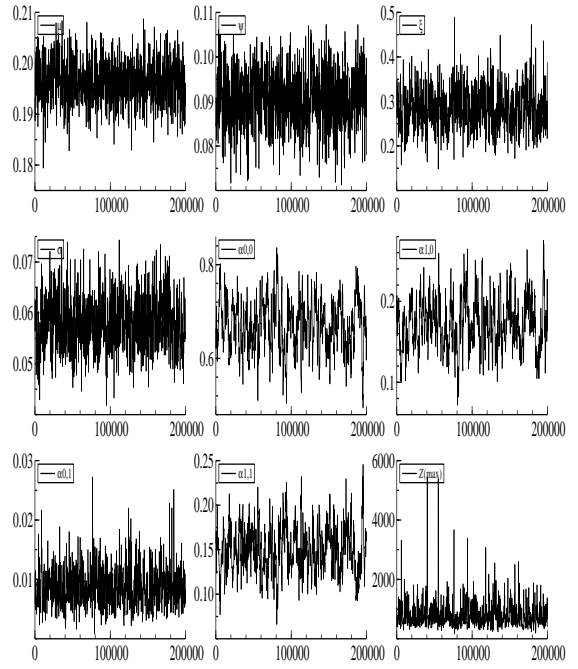

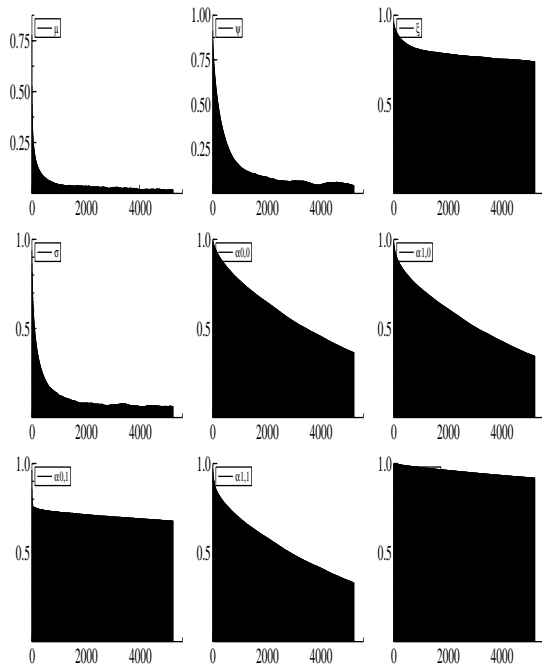Figure 8: Sample paths for simulation data (our sampler).

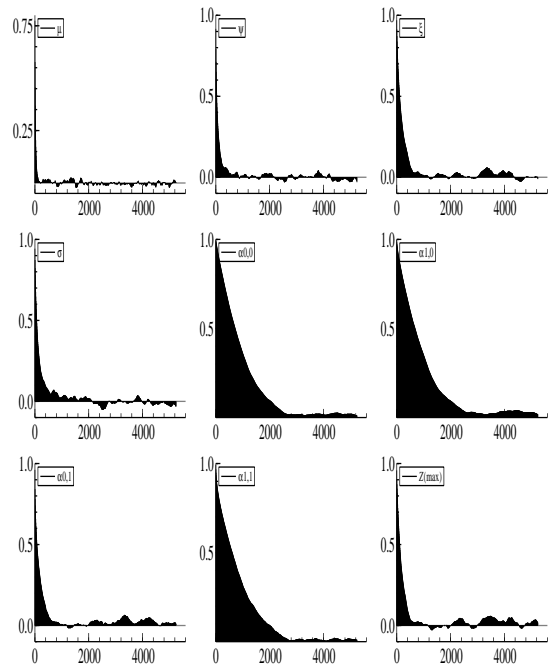Figure 9: Sample autocorrelations for simulation data (the benchmark sampler).

Figure 10: Sample autocorrelations for simulation data (our sampler).

18

5 Application to stock returns data

5.1 Data

In this section, we applied our proposed method to extreme negative returns of high frequency data on the Tokyo Stock Price Index (TOPIX). Currently, increasing numbers of researchers and practitioners are interested in high-frequency financial data to analyse the market dynamic structure. For example, day traders pay much attention to high-frequency data such as 1-min, 3-min, 5-min and 10-min charts. Therefore, the fluctuation in these charts is an important risk factor that greatly affects them.

Moreover, analyzing intraday data is thought to be crucial to maintain stable financial markets. A huge stock market disruption occurred on May 6, 2010, which was referred to as the flash crash. In the crash, the Dow Jones industrial average fell nearly 1,000 points in less than a half-hour. To prevent a recurrence of the flash crash, the Securities and Exchange Commission (SEC) approved new trading rules, known as circuit breakers. Under these rules, trading of any S&P500 stock that rises or falls 10% or more in a five minute period would be halted for five minutes, and initially, these new rules will be tested during a six-month pilot period through December 10, 2010. These facts clearly show the importance of risk analysis using high-frequency data such as 5-min returns to stabilize the equity markets.

We chose 5-min frequency data and analyzed the dynamic structure of the daily minima of 5-min stock return data. Our data were obtained from the Nikkei NEEDS MT tick data, which records the price every minute. The Tokyo Stock Exchange (TSE) is open from 9:00-11:00 (morning session) and 12:30-15:00 (afternoon session) on usual trading days and only for the morning session on the first and last trading days in every year.

Mean	Stdev.	Skewness	Kurtosis	Max.	Min.
0.2780	0.1719	2.6689	14.582	1.6947	0.0423

Table 6: Summary statistics for the TOPIX minimum data (multiplied by -1, $n = 3,321$).

The original sample period was from April 1, 1996 through September 30, 2009, and we used the minimum of 5-min returns in one day as an extreme value. We used a log difference to compute the return (multiplied by 100), excluding the overnight and lunch time intervals, and obtained 24 and 30 5-min intraday returns in the morning and afternoon sessions, respectively. This operation led to a series of 3,321 daily minimum observations. For estimation, we used the minimum series multiplied by -1. Table 6 summarizes the descriptive statistics, and Figure 11 plots the minima time series. The skewness is positive,

and the kurtosis is larger than that of the normal distribution, which implies a longer right tail and fatter tails.
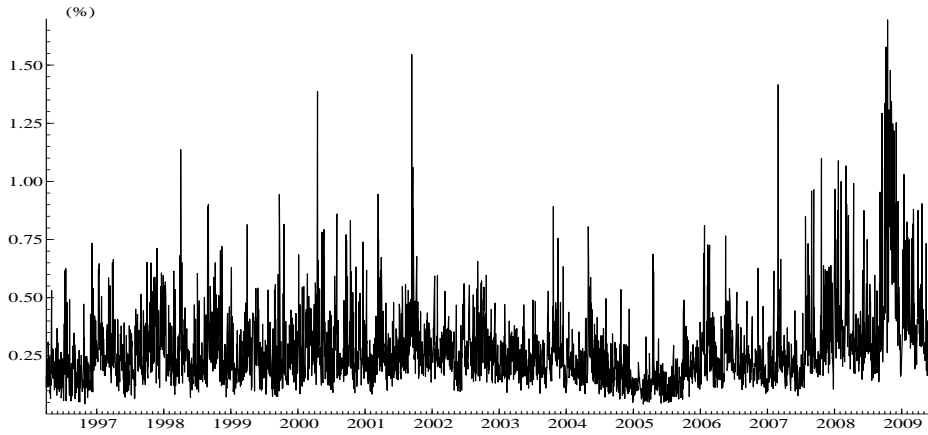

Figure 11: Minimum return data for the TOPIX (multiplied by -1, 1996/April – 2009/September).

5.2 Estimation results

We estimated the GEV-M3 model for TOPIX minima data. The prior specification and the iteration size are the same as in the simulation study.
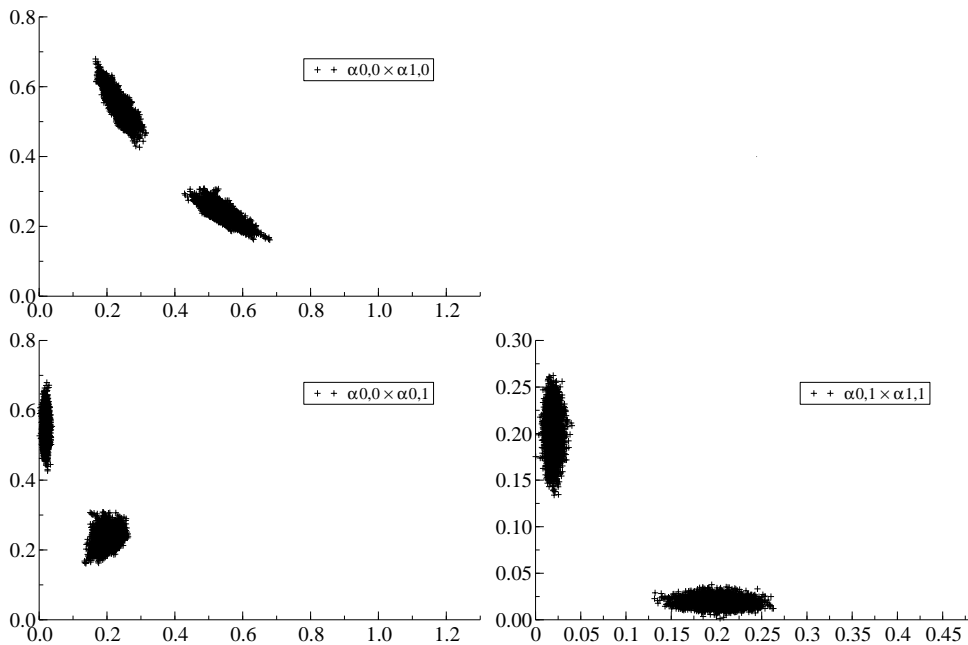

Figure 12: Scatterplot of MCMC draws α from the random permutation sampler for the TOPIX minimum return data (x-label $\times$ y-label).

20

Parameter	GEV	GEV-M3
μ	0.201 (0.001)	0.203 (0.002)
	[0.198, 0.205]	[0.199, 0.207]
	37.2	58.1
ψ	0.079 (0.003)	0.073 (0.003)
	[0.073, 0.085]	[0.066, 0.079]
	128.9	238.1
ξ	0.294 (0.025)	0.382 (0.032)
	[0.247, 0.344]	[0.321, 0.450]
	301.5	596.4
σ	0.041 (0.003)	0.050 (0.002)
	[0.036, 0.047]	[0.046, 0.055]
	172.7	483.8
$\alpha_{0,0}$		0.544 (0.034)
		[0.478, 0.615]
		1918.7
$\alpha_{1,0}$		0.236 (0.024)
		[0.189, 0.291]
		2367.7
$\alpha_{0,1}$		0.019 (0.004)
		[0.011, 0.028]
		445.2
$\alpha_{1,1}$		0.199 (0.018)
		[0.163, 0.235]
		1634.3

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 7: Estimation result of the GEV models for the TOPIX minimum data.

Figures 12 shows the scatter plots of MCMC draws from the random permutation sampler. From this figure, we set the suitable constraint $\alpha_{0,0} > \alpha_{1,0}$ for the GEV-M3 model.

Table 7 reports the estimation result of GEV and GEV-M3 models. Regarding the posterior means for the parameters in the GEV distribution, μ and ψ were not different between these two models, while ξ was larger in the GEV-M3 model than in simple GEV model, which indicates the tail in GEV-M3 model tends to be heavier. In addition, the posterior mean of σ also was larger in this order, which shows that the idiosyncratic error tends to be larger. The posterior means of ξ in both two models were estimated to be positive, and the 95% credible intervals did not contain zero, which indicated that corresponding GEV distributions belong to the Fréchet type and have heavier tails than other types.

Next, we will discuss the estimation results concerning time dependence. For the $(\alpha_{0,0}, \alpha_{0,1})$ corresponded to the index $l = 0$, and the posterior mean of $\alpha_{0,0}$ was much larger than that of $\alpha_{0,1}$. This result implies that one part of the M3 process has a weak time-dependent structure. On the other hand, the posterior means of $\alpha_{1,0}$ and $\alpha_{1,1}$ were relatively similar. This indicates that the impact of the current shock appeared to remain

in the next period and that the other part of the M3 process has strong time dependence.

To measure the strength of dependence of the process, the extremal index θ was suggested to be a key parameter (e.g., Coles (2001) and Leadbetter et al. (1983)). Loosely, the extremal index can be interpreted as the inverse of the mean cluster size in the point process of exceedance times over a high threshold. For independent series, the extremal index $\theta = 1$. The extremal index of the M3 process is given by $\theta = \sum_l \max_k \alpha_{l,k}$ (see e.g., Smith and Weissman (1996)). For the TOPIX data, the estimation result is reported in Table 8. The posterior mean was around 0.8, apart from 1, and considering the 95% interval, this result shows that time dependence certainly exists in the TOPIX minimum return data.

Finally, Figures 13, 14 and 15 show the sample paths, the sample autocorrelations, and the posterior densities for the GEV-M3 model, respectively. The MCMC results show that the Markov chains mixed well.

Parameter	Mean	Stdev.	95% interval
θ_{M3}	0.780	0.018	$[0.746, 0.817]$

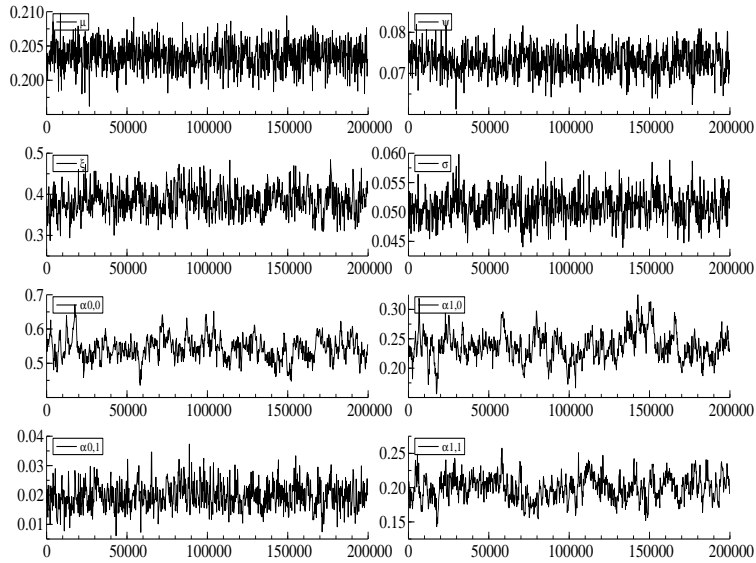Table 8: Estimated extremal index for the TOPIX minimum return data.

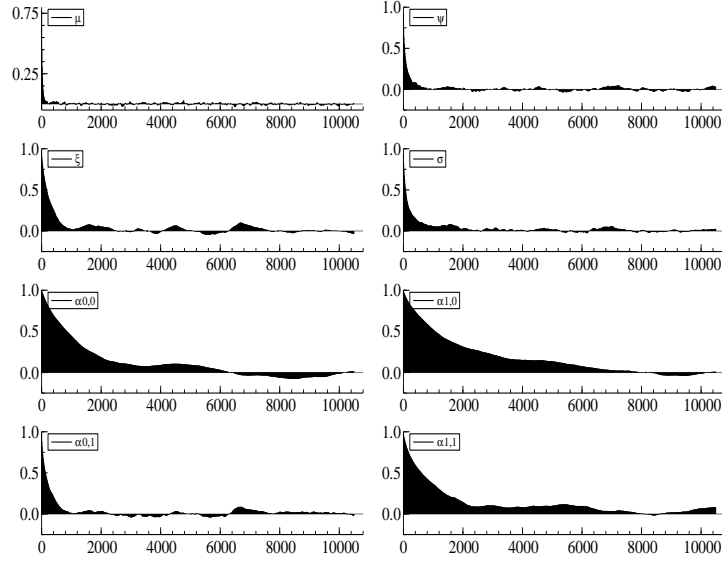Figure 13: Sample paths for the TOPIX minimum return data.

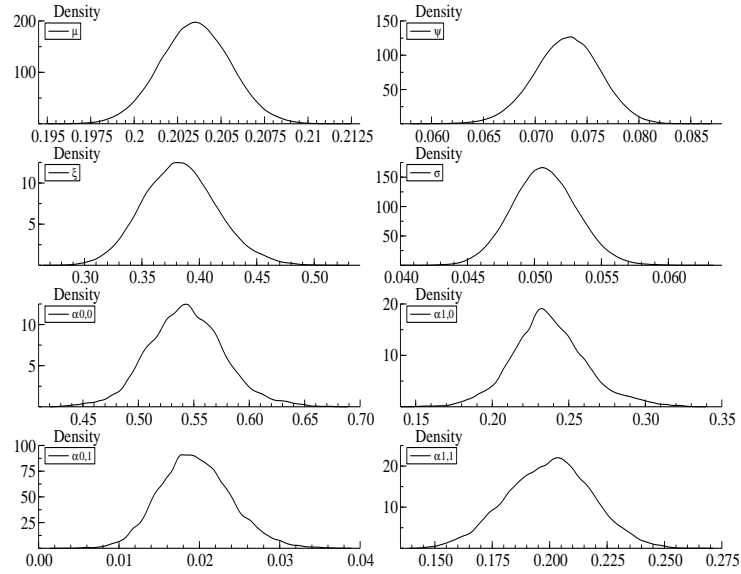Figure 14: Sample autocorrelations for the TOPIX minimum return data.

Figure 15: Posterior densities for the TOPIX minimum return data.

5.3 Model comparison using marginal likelihoods

In this subsection, we compare GEV and GEV-M3 models using marginal likelihood. The marginal likelihood is defined as the integral of the likelihood with respect to the prior density of the parameter. Following Chib (1995), the basic marginal likelihood identity is

satisfied for any $\vartheta \in \Theta$ (Θ:parameter space):

$$m(y) = \frac{f(y|\vartheta)\pi(\vartheta)}{\pi(\vartheta|y)}$$

where $m(y)$ is marginal likelihood, $\pi(\vartheta)$ is a prior density, $f(\vartheta|y)$ is a likelihood and $\pi(\vartheta|y)$ is a posterior density. Here if we take the log, then

$$\log m(y) = \log f(y|\vartheta) + \log \pi(\vartheta) - \log \pi(\vartheta|y).$$

If we evaluate the prior, posterior and likelihood at the fix point $\vartheta = \vartheta^*$, we can estimate the marginal likelihood $m(y)$. Theoretically, any $\vartheta^* \in \Theta$ can be used, but we typically use the posterior mean or mode for stable estimation of $m(y)$. The prior density is easily evaluated, and the likelihood is calculated by the filtering method discussed in Subsection 5.4. For the posterior ordinate, we used a previously described method (Chib (1995) and Chib and Jeliazkov (2001)), using samples from the reduced form of the MCMC algorithm.

In our study, the likelihood was estimated using 10,000 particles, and 10 replications of the filter were implemented to compute the standard error of the likelihood. The posterior ordinate was estimated by the reduced MCMC runs using 10,000 and 50,000 draws for GEV and GEV-M3 respectively.

Model	GEV	GEV-M3
Likelihood ordinate	2176.18	2279.93
(S.E.)	(0.66)	(0.93)
Prior ordinate	-8.47	-3.65
Posterior ordinate	21.05	35.43
(S.E.)	(0.17)	(0.31)
Marginal likelihood	2146.66	2240.84
(S.E.)	(0.68)	(0.98)

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

Table 9: Estimated marginal likelihood for the TOPIX minimum return data.

Table 9 reports the estimation result of the marginal likelihood. Although the marginal likelihood of the GEV-M3 model was more penalized by the prior and the posterior ordinates due to additional parameters compared with that of GEV model, the result shows that the GEV-M3 model outperforms the simple GEV model. Therefore, the importance of capturing time dependence in a financial time series is clear.

5.4 Efficiency comparison of proposed particle filters

Using the daily minima stock returns, this subsection compares our proposed particle filter (denoted by 'New') with other filters: simple particle filter (PF) and auxiliary particle filter (APF, Pitt and Shephard (1999)). For each filter, we used $R = 500, 1000, 5000$, and 10000 particles to compute the likelihood and its standard error was obtained based on 10 replications of the filter.

Table 10 shows the estimation result, where we found that our filter produces the most accurate estimates among all filters for both models. In addition, our estimates were found to be very stable, since the differences in estimates among $R = 500, 1000, 5000$, and 10000 cases were the smallest and each standard error was less than those of the other filters.

Since the simple particle filter (PF) uses unit Fréchet as an importance function, it fails to generate sufficient particles from the tail of the unit Fréchet distribution, and hence most weights $w_{t+1}^i = f(y_{t+1}|x_{t+1}^i, \vartheta)$, $i = 1, \ldots, I$ are likely to be small especially when R is small. This could lead to inaccurate estimates for small values of R in PF. In addition, the standard errors were very large, further indicating its lower stability compared with our filter.

Concerning APF, the estimates were incorrectly much smaller than those in New and PF. To determine how these incorrect estimates were obtained, we considered the weights for APF given by

$$w_{t+1}^i = \frac{f(y_{t+1}|x_{t+1}^i, \vartheta)f(z_{t+1}^i)\hat{f}(z_{t+1-K:t}^i|y_{1:t}, \vartheta)}{f(z_{t+1}^i)h(z_{t+1-K:t}^i|y_{1:t}, \vartheta)} = \frac{f(y_{t+1}|x_{t+1}^i, \vartheta)\hat{f}(z_{t+1-K:t}^i|y_{1:t}, \vartheta)}{h(z_{t+1-K:t}^i|y_{1:t}, \vartheta)},$$

where

$$h(z_{t+1-K:t}^i|y_{1:t}, \vartheta) = \frac{f(y_{t+1}|\nu_{t+1}^i, \vartheta)\hat{f}(z_{t+1-K:t}^i|y_{1:t}, \vartheta)}{\sum_{j=1}^M f(y_{t+1}|\nu_{t+1}^j, \vartheta)\hat{f}(z_{t+1-K:t}^j|y_{1:t}, \vartheta)},$$

and ν_{t+1}^i is equal to x_{t+1}^i except $z_{l,t+1}$ in x_{t+1}^i is replaced by the median of unit Fréchet distribution ($= 1/\log 2$) as a likely value for $l = 0, 1, \ldots, L$. When y_{t+1} is large, ν_{t+1} fails to approximate x_{t+1} and most particles generated from the importance function $h(z_{t+1-K:t}^i|y_{1:t}, \vartheta)$ have almost zero weights. Additionally, only a small number of particles have large weights. As a result, such particles with high $h(z_{t+1-K:t}^i|y_{1:t}, \vartheta)$ values are likely to be sampled in the first step. In the second step, z_{t+1} was generated from unit Fréchet as PF. Therefore, APF leads to a case with large $h(z_{t+1-K:t}^i|y_{1:t}, \vartheta)$ and small $f(y_{t+1}|x_{t+1}^i, \vartheta)$ values' for most particles, which could cause much smaller estimates as shown in Table 10.

Based on these findings, it is clear that our filtering method outperforms other filters.

	New	PF	APF
$R = 500$	2273.81	1931.08	-270.97
	(3.18)	(202.73)	(83.97)
$R = 1,000$	2276.48	2125.75	11.87
	(2.50)	(133.62)	(80.52)
$R = 5,000$	2279.65	2271.15	441.9
	(0.93)	(4.98)	(23.92)
$R = 10,000$	2279.93	2275.95	571.95
	(0.93)	(3.24)	(28.3)

*All values are in natural log scale. Standard errors are in parentheses. R denotes the number of particles.

Table 10: Estimated log-likelihoods for the TOPIX minimum data using three particle filter methods; New (proposed filter), PF (simple particle filter) and APF (Auxiliary particle filter).

6   Conclusion

This paper discusses Bayesian estimation for time dependent extreme value models based on M3 processes. We proposed an improved MCMC algorithm using a mixture sampler, which approximated Gumbel density by a ten-component mixture of normal. A simulation study showed that our MCMC algorithm works successfully and outperforms a benchmark method. In addition, we proposed an associated particle filter method that generates particles directly based on the data information and demonstrated that our filtering method is superior to those reported previously. In application to high-frequency TOPIX minimum returns data, the parameter estimates showed that the daily series of minimum returns certainly has time dependence. Moreover, a model comparison using marginal likelihood indicated that the GEV-M3 model outperforms the simple GEV model.

A Details on MCMC Estimation

A.1 GEV-M3 Model

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

In Step 1 of Algorithm 1, the conditional posterior density of $\lambda = (\mu, \psi, \xi)'$ is given by $\pi(\lambda|\sigma^2, x, y) \propto \pi(\lambda)f(y|\lambda, \sigma^2, x)$, where f is the conditional likelihood of the observation equation (4). To sample from this conditional posterior distribution, we utilized the Metropolis-Hastings (M-H) algorithm using a normal proposal density as follows. First, we determined $\hat{\lambda} = (\hat{\mu}, \hat{\psi}, \hat{\xi})'$ which maximizes (or approximately maximizes) the conditional posterior density. Next, we generated a candidate λ^* from a normal distribution truncated over the region $R = \{\lambda : \psi > 0\}$, $\mathcal{TN}_R(\lambda_*, \Sigma_*)$, where

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

and accept with probability

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

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^* was rejected, we took λ as a next sample.

A.1.2 Generation of s

We drew a sample $s_{l,t}$ from its discrete conditional posterior distribution with a probability mass function,

$$\pi(s_{l,t} = i|w) \propto p_i \frac{1}{v_i} \exp \left\{ -\frac{1}{2v_i^2}(w_{l,t} - m_i)^2 \right\}$$

for $l = 0, \ldots, L$, $i = 1, \ldots, S$, and $t = 0, \ldots, n$.

A.1.3 Sampling state variables

We sampled $w_{l,t} = \log z_{l,t}$ from the conditional posterior distribution. Taking the relation between w and x into account, $w_{l,t}$ is contained as an element in $x_t, \ldots, x_{t+K}$. We defined $\hat{x}_{l,t+k}$, which is a maximum of the components in x_{t+k} except the element of $w_{l,t}$, as

$$\hat{x}_{l,t+k} = \max_{l',k'}\{\alpha_{l',k'} \exp(w_{l',t+k-k'})| \ 0 \leq l' \leq L, 0 \leq k' \leq K, (l',k') \neq (l,k)\}, \ 0 \leq k \leq K.$$

$\alpha_{l,k} \exp(w_{l,t}) \geq \hat{x}_{l,t+k}$ means $\alpha_{l,k} \exp(w_{l,t}) = x_{t+k}$, and we used information about y_{t+k} to construct proposal distributions. Otherwise, since $\alpha_{l,k} \exp(w_{l,t}) < \hat{x}_{l,t+k}$ indicates that $w_{l,t}$ has no influence on x_{t+k}, the likelihood function $f(y_{t+k}|x_{t+k}, \vartheta)$ can be considered a constant term in the posterior representation of $w_{l,t}$. Also $\alpha_{l,k} = 0$ implies that $f(y_{t+k}|x_{t+k}, \vartheta)$ can be treated as a

constant term. Therefore, when $\alpha_{l,k} = 0$ for all k, we generated samples simply from Gumbel(0,1) distribution. When $\alpha_{l,k} > 0$ for some k, the inequality $\alpha_{l,k} \exp(w_{l,t}) \geq \hat{x}_{l,t+k}$ is equivalent to $w_{l,t} \geq \log(\hat{x}_{l,t+k}/\alpha_{l,k})$, and we allowed M to denote the number of such $\alpha_{l,k}$ values. We defined $x^*_{l,t+k} = \log(\hat{x}_{l,t+k}/\alpha_{l,k})$ as a boundary point for $w_{l,t}$ to be included in the likelihood. By arranging them in ascending order such that $x^*_{l,t+k_1} \leq \ldots \leq x^*_{l,t+k_m} \leq \ldots \leq x^*_{l,t+k_M}$, we divided the real line as follows:

$$
\begin{aligned}
R_0 &= \{w_{l,t} | w_{l,t} < x^*_{l,t+k_1}\} \\
R_j &= \{w_{l,t} | x^*_{l,t+k_j} \leq w_{l,t} < x^*_{l,t+k_{j+1}}\}, \ j = 1, \ldots, M-1, \\
R_M &= \{w_{l,t} | x^*_{l,t+k_M} \leq w_{l,t}\}.
\end{aligned}
$$

In the interval R_0, $\alpha_{l,k} \exp(w_{l,t}) < \hat{x}_{l,t+k}$ for all k and the conditional posterior density is given by

$$
\pi(w_{l,t}|y, \vartheta, w/w_{l,t}, s) \propto \exp\left\{-\frac{1}{2v^2_{s_{l,t}}}\left(w_{l,t} - m_{s_{l,t}}\right)^2\right\}. \tag{15}
$$

Then, we generated samples from $\mathcal{TN}_{R_0}(\bar{w}_{l,t}, \bar{\sigma}^2_{l,t})$ where $\bar{w}_{l,t} = m_{s_{l,t}}$ and $\bar{\sigma}^2_{l,t} = v^2_{s_{l,t}}$.

In other intervals, R_i for $i = 1, \ldots, M$, the conditional posterior density of $w_{l,t}$ is

$$
\begin{aligned}
\pi(w_{l,t}|y, \vartheta, w/w_{l,t}, s) &\propto \exp\left\{-\frac{1}{2v^2_{s_{l,t}}}\left(w_{l,t} - m_{s_{l,t}}\right)^2\right\} \\
&\times \exp\left[-\frac{1}{2\sigma^2}\sum_{m=1}^{i}\left\{y_{t+k_m} - \mu - \psi\frac{\alpha^\xi_{l,k_m}\exp(\xi w_{l,t}) - 1}{\xi}\right\}^2\right].
\end{aligned}
$$

To construct a normal proposal density, we considered the second-order Taylor expansion of the logarithm of the likelihood (excluding the constant term)

$$
h(w_{l,t}) \equiv -\frac{1}{2\sigma^2}\sum_{m=1}^{i}\left\{y_{t+k_m} - \mu - \psi\frac{\alpha^\xi_{l,k_m}\exp(\xi w_{l,t}) - 1}{\xi}\right\}^2,
$$

around the mode $\hat{w}_{l,t}$. We allowed $h'(\hat{w}_{l,t})$ and $h''(\hat{w}_{l,t})$ to denote the first and the second derivative of $h(w_{l,t})$ evaluated at $w_{l,t} = \hat{w}_{l,t}$, respectively. Then,

$$
\begin{aligned}
h(w_{l,t}) &\approx h(\hat{w}_{l,t}) + h'(\hat{w}_{l,t})(w_{l,t} - \hat{w}_{l,t}) + \frac{1}{2}h''(\hat{w}_{l,t})(w_{l,t} - \hat{w}_{l,t})^2 \\
&= \frac{1}{2}h''(\hat{w}_{l,t})\left\{w_{l,t} - \left(\hat{w}_{l,t} - \frac{h'(\hat{w}_{l,t})}{h''(\hat{w}_{l,t})}\right)\right\}^2 + \text{const.} \\
&= -\frac{(w^*_{l,t} - w_{l,t})^2}{2\sigma^{*2}_{l,t}} + \text{const.},
\end{aligned}
$$

where $\sigma^{*2}_{l,t} = -\{h''(\hat{w}_{l,t})\}^{-1}$ and $w^*_{l,t} = \hat{w}_{l,t} + \sigma^{*2}_{l,t}h'(\hat{w}_{l,t})$. Based on this approximation, we used

the following proposal density q_i in R_i.

$$q_i(w_{l,t}|y,\vartheta,w/w_{l,t},s) \quad \propto \quad \exp\left\{-\frac{(w_{l,t}^*-w_{l,t})^2}{2\sigma_{l,t}^{*\,2}}\right\}\exp\left\{-\frac{(w_{l,t}-m_{s_{l,t}})^2}{2v_{s_{l,t}}^2}\right\}$$

$$\propto \quad \exp\left\{-\frac{(w_{l,t}-\bar{w}_{l,t})^2}{2\bar{\sigma}_{l,t}^2}\right\}, \tag{16}$$

where

$$\bar{\sigma}_{l,t}^2 = \left(\sigma_{l,t}^{*\,-2}+v_{s_{l,t}}^{-2}\right)^{-1}, \quad \bar{w}_{l,t}=\bar{\sigma}_{l,t}^2\left(\frac{w_{l,t}^*}{\sigma_{l,t}^{*\,2}}+\frac{m_{s_{l,t}}}{v_{s_{l,t}}^2}\right).$$

If $\bar{w}_{l,t}$ was smaller than the lower bound of R_i, that is, $x_{l,t+k_i}^*$, or larger than the upper bound of R_i, $x_{l,t+k_{i+1}}^*$, then we reset $\bar{w}_{l,t}=x_{l,t+k_i}^*$ or $\bar{w}_{l,t}=x_{l,t+k_{i+1}}^*$ respectively. Then, we approximated the posterior density in R_i using truncated normal $N_{R_i}(\bar{w}_{l,t},\bar{\sigma}_{l,t}^2)$.

Finally, we constructed the proposal distribution q combining proposals in each interval R_i as follows:

$$q(w_{l,t}|y,\vartheta,w/w_{l,t},s)=\sum_{i=0}^M \frac{p_i}{\sum_{j=0}^M p_j}q_i(w_{l,t}|y,\vartheta,w/w_{l,t},s), \tag{17}$$

where q_i is the truncated normal proposal in region R_i and

$$p_i=\int_{R_i}\frac{1}{\sqrt{2\pi}\bar{\sigma}_{l,t}}\exp\left\{-\frac{(w_{l,t}-\bar{w}_{l,t})^2}{2\bar{\sigma}_{l,t}^2}\right\}dw_{l,t}.$$

A.2   Associated efficient particle filter

Subsection 3.3 discusses our proposed particle filter based on the auxiliary mixture sampler. We constructed $g(z_{l^*,t+1}|y_{t+1},z_{l,t+1},z_{t+1-K:t},s_{l^*,t+1},\vartheta)$ in (12) using the same mixture approximation method as in Appendix A.1.3. Since the data information $y_{1:t+1}$ is available up to time $t+1$, $w_{l,t+1}$ could appear only through x_{t+1}. Based on $x_{l,t+1}^*$, we divided the domain of $w_{l,t}$ into two intervals, $R_0=\{w_{l,t+1}|w_{l,t+1}<x_{l,t+1}^*\}$ and $R_1=\{w_{l,t+1}|w_{l,t+1}\geq x_{l,t+1}^*\}$ where the proposals are given by (15) and (16) respectively. Finally, we used a normal mixture distribution (17) with the two intervals as an importance function.

References

Beirlant, J., Y. Goefebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

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 E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory 12*, 409–431.

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.

Davis, R. and S. Resnick (1989). Basic properties and prediction of max-arma processes. *Advances in applied probability 21*, 781–803.

Davis, R. and S. Resnick (1993). Prediction of stationary max-stable processes. *Annals of Applied Probability 3*, 497–525.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* Berlin: Springer.

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.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo methods in practice.* New York: Springer.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events: for Insurance and Finance.* New York: Springer.

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.

Früwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*, 194–209.

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.

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.

Hall, P., L. Peng, and Q. Yao (2002). Moving-maximum models for extrema of time series. *Journal of Statistical Planning and Inference 103*, 51–63.

Heffernan, J. E., J. Tawn, and Z. Zhang (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes. *Extremes 10*, 57–82.

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.

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

Leadbetter, K. R., G. Lindgren, and H. Rootzén (1983). *Extremes and related properties of random sequences and series*. New York.

Martins, A. and H. Ferreira (2005). The multivariate extremal index and the dependence structure of a multivariate extreme value distribution. *TEST 14*, 433–448.

Nakajima, J., T. Kunihama, Y. Omori, and S. Früwirth-Schnatter (2009). Generalized extreme value distribution with time-dependence using the ar and ma models in state space form. Manuscript.

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

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

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.

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.

Zhang, Z. (2002). Multivariate extremes, max-stable process estimation and dynamic financial modeling. PhD dissertation, Department of Statistics, University of North Carolina, Chapel Hill.

Zhang, Z. (2009). On approximating max-stable processes and constructing extremal copula functions. *Statistical Inference for Stochastic Processes 12*, 89–114.

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

Zhang, Z. and R. L. Smith (2010). On the estimation and application of max-stable processes. *Journal of Statistical Planning and Inference 140*, 1135–1153.