

## An extended sparse max-linear moving model with application to high-frequency financial data

Timothy Idowu & Zhengjun Zhang

To cite this article: Timothy Idowu & Zhengjun Zhang (2017) An extended sparse max-linear moving model with application to high-frequency financial data, *Statistical Theory and Related Fields*, 1:1, 92-111, DOI: [10.1080/24754269.2017.1346852](https://doi.org/10.1080/24754269.2017.1346852)

To link to this article: <https://doi.org/10.1080/24754269.2017.1346852>



Published online: 28 Jul 2017.



Submit your article to this journal [↗](#)



Article views: 122



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)



# An extended sparse max-linear moving model with application to high-frequency financial data

Timothy Idowu and Zhengjun Zhang

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

## ABSTRACT

There continues to be unfading interest in developing parametric max-stable processes for modelling tail dependencies and clustered extremes in time series data. However, this comes with some difficulties mainly due to the lack of models that fit data directly without transforming the data and the barriers in estimating a significant number of parameters in the existing models. In this work, we study the use of the sparse maxima of moving maxima (M3) process. After introducing random effects and hidden Fréchet type shocks into the process, we get an extended max-linear model. The extended model then enables us to model cases of tail dependence or independence depending on parameter values. We present some unique properties including mirroring the dependence structure in real data, dealing with the undesirable signature patterns found in most parametric max-stable processes, and being directly applicable to real data. A Bayesian inference approach is developed for the proposed model, and it is applied to simulated and real data.

## ARTICLE HISTORY

Received 2 April 2017  
Revised 14 June 2017  
Accepted 22 June 2017

## KEYWORDS

Extreme value theory; max-stable processes; time series; Bayesian inference; max-linear models; high-frequency financial data

## 1. Introduction

With financial crisis having hit the global market at different times, the requirement on risk analysts to better estimate risk has become stronger; hence, accurately estimating risk measures such as value at risk (VaR) or expected shortfall (ES) is increasingly important. It is easily noticeable that when an extreme price drop or rise happens in financial markets, there is a high likelihood for such event to reoccur or continue within a short period; therefore, studying the dependence structure of risk variables especially within tail regions is important.

The pioneering work of Box and Jenkins (1970) introduced the autoregressive and moving average (ARMA) model to model dependence structure in time series data. On the other hand, the works of Fisher, Tippett, and Gnedenko (see Fisher & Tippett, 1928; Gnedenko, 1943) on the maximum of the sequence of independent and identically distributed random variables pioneered the study of extremes. The role of their theorem with regard to maxima is analogous to that of the central limit theorem with regard to sample means. In the study of extremes and tail dependence, these pioneering works have been followed by several other proposals with varying objectives.

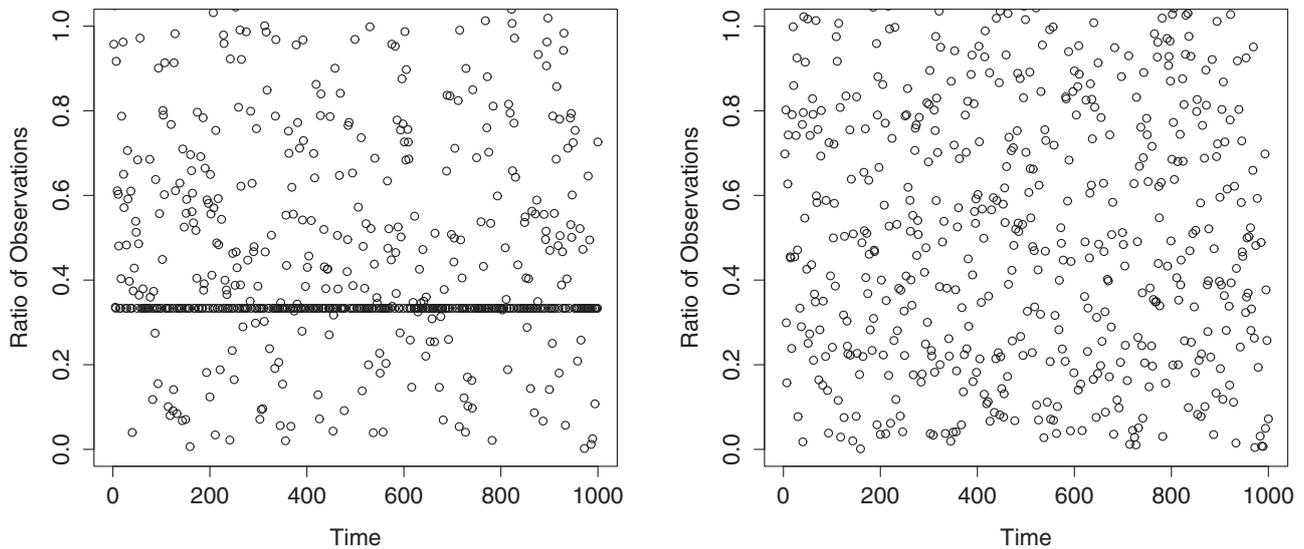
As attention to risk analysis continues to increase, the extreme value theory which has been well applied in diverse fields of studies like meteorology, geology and finance remains a useful tool in describing the properties of the usually heavy-tailed risk variables. Looking towards the extreme value theory, we can find some parametric max-stable processes adaptable to various

dependence structures over time that have been proposed in different forms. Among them is the maxima of moving maxima (M3) process with the following representation proposed by Deheuvels (1983),

$$X_t = \max_{l \geq 1} \max_{-\infty < k < \infty} \alpha_{lk} Z_{l,t-k}, \quad -\infty < t < \infty, \quad (1)$$

where  $\{\alpha_{lk}, -\infty < k < \infty, l \geq 1\}$  is an array of non-negative parameters such that  $\sum_{lk} \alpha_{lk} = 1$  and  $\{Z_{lt}, -\infty < t < \infty, l \geq 1\}$  is an array of *iid* unit Fréchet random variables i.e.  $F_Z(z) = \exp(-1/z)$  for  $z > 0$ . Some other examples are the max-autoregressive moving average (MARMA) process in Davis and Resnick (1989), the extension of the M3 process to the multivariate (maxima of moving maxima (M4) process) case in Smith and Weissman (1996) and recently the sparse moving maxima model (SM4R) in Tang, Shao, and Zhang (2013).

Most of the aforementioned moving maxima models have some shortcomings we intend to address. First is the difficulty in estimating parameters as a result of a vast number of parameters and sometimes an infinite number of them (e.g. the M3 process). Also, the difficulty can be attributed to the often complex nature of the distribution functions of the models and their structures. These challenges make models with a parsimonious number of parameters more attractive. Second is the difficulty in suiting the lag tail-dependence structures in real data (see Zhang, 2005). Third is due to the lack of the ability of mirroring real data because of inherent signature patterns, that is, ratios of observations having constant values as shown in Figure 1



**Figure 1.** Left side: Signature pattern of MM(1) process with  $X_{t+1}/X_t = (1 - \alpha)/\alpha = 1/3$ . It is shown in Zhang and Smith (2010) that signature patterns will occur infinitely many times in a moving maxima processes. Right side: The loading constant  $\alpha$  is replaced with a uniform random variable on the interval  $[0, 1]$  to remove the signature pattern.

illustrated using the MM(1) process<sup>1</sup>. To overcome this, we add random effects to the loading constants of our sparse matrix as proposed in Tang et al. (2013). The usefulness of these random effects can be seen on the right-hand side of Figure 1. Finally, the existing models are often not directly applicable to fit real data. Rather, they are often applied to fit transformed data, for example, residuals extracted from ARMA-GARCH filters utilised by McNeil and Frey (2000) among others. In this work, we aim to tackle these problems by modifying some previously proposed models.

We employ sparsity in the dependence structure of our model to better mirror real data and aid estimation of parameters, while random effects are employed to deal with signature patterns. Another advantage of our proposed model is that it is directly applicable to data without the usual requirement for complex transformations in the max-stable process framework. To tackle the estimation problem, we develop a Bayesian inference approach.

In the subsequent sections, we proceed as follows. Section 2 discusses our proposed max-linear model and estimation approach. Simulation examples are presented in Section 3. We further apply our model to real data of currency prices in Section 4 and conclude in Section 5. Proofs and graphs are provided in Section 6.

## 2. Model

In this section, we propose a model for modelling tail (in)dependence and clustering of extremes in univariate financial time series data. We apply Bayesian inference for parameter estimation using the Metropolis–Hasting algorithm from Metropolis, Rosenbluth, Rosenbluth,

Teller, and Teller (1953) and Hastings (1970) in tandem with the Gibbs sampling technique by Geman and Geman (1984) in the so-called hybrid MCMC (see Robert & Casella, 2004).

### 2.1. Model set-up

Our aim is to have a model whose implied dependence structure well describes what is in real data, its parameters can be well estimated, and it can be fitted to data directly so that interpretation from inference can be easier.

The literature (Smith & Weissman, 1996) showed that extremal properties of stationary time series could be studied using limiting max-stable processes. They also showed that the maxima of moving maxima process could be used to approximate a broad class of max-stable processes. The univariate M3 process is, therefore, a useful model for studying extremes and clustering in stationary time series. Below is a definition of a max-stable process which considers unit Fréchet marginals. As for other extreme value settings, note that transformation can be made back and forth between the extreme value type distributions.

**Definition 2.1:** A random process  $\{X_t, t = 1, 2, \dots\}$  with unit Fréchet marginals such that

$$P^n(X_t \leq nx_t, 1 \leq t \leq s) = P(X_t \leq x_t, 1 \leq t \leq s) \quad (2)$$

for any  $s \geq 1$  and  $n \geq 1$  is a max-stable process.

In our model set-up, we add sparsity and random effects to the loading matrix of the M3 process. Note that Zhang and Zhu (2016) proposed a similar model. Here, we extend the model within the specification of

random shocks. We, therefore, start with the following form:

$$V_t = \max [\mathbf{A}_t \cdot \mathbf{Z}_t], \quad -\infty < t < \infty, \quad (3)$$

where  $\mathbf{A}_t \cdot \mathbf{Z}_t$  is a component-wise product with

$$\mathbf{Z}_t = \begin{pmatrix} Z_{1t} & Z_{1,t-1} & \cdots & Z_{1,t-L} \\ Z_{2t} & Z_{2,t-1} & \cdots & Z_{2,t-L} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{Lt} & Z_{L,t-1} & \cdots & Z_{L,t-L} \end{pmatrix},$$

and

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1 U_{1t} & \alpha_1 (1 - U_{1t}) & 0 & \cdots & 0 \\ \alpha_2 U_{2t} & 0 & \alpha_2 (1 - U_{2t}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_L U_{Lt} & 0 & 0 & \cdots & \alpha_L (1 - U_{Lt}) \end{pmatrix}.$$

We have  $\max [\mathbf{A}_t \cdot \mathbf{Z}_t]$  as the maximum of the components of the matrix.  $\alpha_l \geq 0$  for each  $l = 1, \dots, L$  and  $\sum_l \alpha_l = 1$  (meaning  $V_t$ 's are marginally distributed as unit Fréchet).  $\{Z_{lt}, l = 1, \dots, L, -\infty < t < \infty\}$  are a double sequence of *iid* unit Fréchet random variables.  $\{U_{lt}, l = 1, \dots, L, -\infty < t < \infty\}$  are *iid* random variables on the interval  $[0,1]$  (uniformly distributed in our analysis and examples) independent of  $Z_{lt}$ 's. Here,  $\{\alpha_l\}$  are for scaling of the lag effects while  $\{U_{lt}\}$  are for random effects. These two together determine the dependence structure in  $V_t$ .

The model in (3) which builds on the M3 process specifies temporal dependence. However, there are often transient effects or random shocks (not necessarily unit Fréchet distributed) observable in real time series. For example in financial data, there will be market shocks that come and go away without any continuous impacts. To better mirror real data, we add random shocks by introducing hidden max Fréchet shock random variables (see for example Heffernan, Tawn, & Zhang, 2007). This new addition also determines asymptotic dependence or independence structure of the model (see Proposition 2.2). We therefore have,

$$Y_t = \max (W_t, \max [\mathbf{A}_t \cdot \mathbf{Z}_t]), \quad -\infty < t < \infty. \quad (4)$$

$\{W_t, -\infty < t < \infty\}$  is a sequence of *iid* Fréchet( $\beta, 1, \tau$ ) random variables (with distribution function  $\exp(-(w - \tau)^{-\beta})$  for  $w > \tau$ ) independent of  $Z_{lt}$ 's.  $\beta > 0$  is a shape parameter,  $\tau \in \mathbb{R}$  is a location parameter and the scale parameter is fixed as 1. Based on the above specifications, it should be noted that  $W_t$  and  $V_t$  are independent processes. We also note that the existing models in the literature specified  $\tau = 0$ . To have  $\tau \neq 0$  makes the model more applicable to real data, which also increases the estimation difficulty.

Our model in (4) has two parts, the sparse M3 part and the hidden shock part. When  $\beta > 1$ , the hidden shock is dominated by the sparse M3 process in the

upper tail region. The reverse is true when  $\beta < 1$ , while neither dominates with  $\beta = 1$ . See the asymptotic properties in Proposition 2.2 for more details. As for the location parameter,  $\tau < 0$  means the hidden shock has a reduced impact in the lower tail region compared to the sparse M3 process, and the reverse is the case for  $\tau > 0$ . See Figure A3 for illustration using the density functions.  $\beta$  and  $\tau$  therefore specify the role of  $W_t$ 's as shock variables (or noise terms). An extreme value from  $W_t$  implies that  $Y_t$  will be large while extreme value from  $Z_{lt}$  implies  $Y_t$  and  $Y_{t+l}$  will likely both be large depending on the scaling constants ( $\alpha_l$ 's) and random coefficients ( $U_{lt}$ 's).

It is unlikely, however, to observe data that will be on the same scale as the model in (4); hence, we scale it with a parameter  $C$ . To further fit the shape of observable data more precisely, we add a shape parameter  $\psi$ . These parameterisations are reasonable. For example, the generalised extreme value distribution often applied to real data has scale and shape parameters. Our final model is, therefore, applicable to univariate financial time series data without any complex transformation of the data. Note that a data transformation can make interpretation of results difficult. Finally, we extend our model in (4) in the following manner:

$$X_t = CY_t^{1/\psi}, \quad -\infty < t < \infty, \quad (5)$$

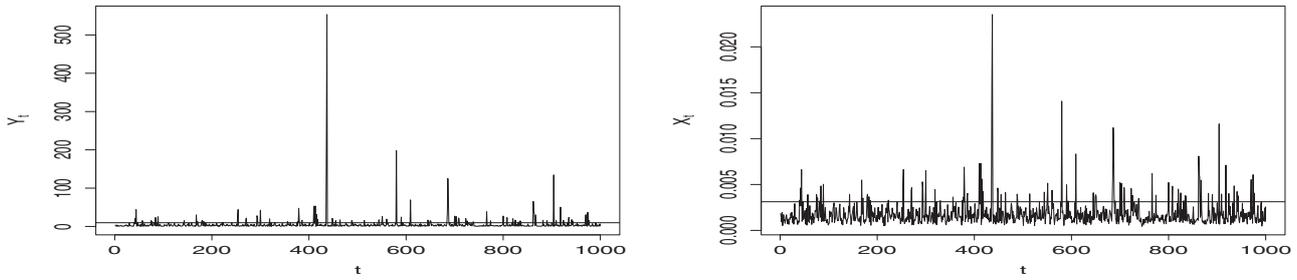
where  $C > 0$  and  $\psi > 0$  are scale and shape parameters, respectively.

Our models (4) and (5) share the same structure of the models proposed in [26] where generalised method of moments (GMM) estimation procedure was utilised. In this work, we introduce an additional location parameter  $\tau$  and develop a Bayesian inference approach which enables us to recover the strengths of random shock effects of  $W_t$  and to estimate the location parameter  $\tau$ .

## 2.2. Model illustration

Two simulation examples of models (4) and (5) are shown in Figure 2. These two examples are from the same  $V_t$  and  $W_t$  latent processes. The parameters chosen here are close to the estimated parameters in our real data example in Section 4. The scaling parameters in  $\alpha$  ( $L$  set to 3), which determined the dependence structure in  $V_t$  and hence in  $Y_t$  and  $X_t$ , are set to  $(\alpha_1, \alpha_2, \alpha_3) = (0.5, 0.3, 0.2)$ .

To highlight the roles of the two latent processes, using the same latent processes as the simulations in Figure 2, we present more details in Figure 3. On the left top-side we have the first 50 realisations. The dots are values of the latent process  $V_t$  when it equals  $Y_t$ . On the top right, we have the 901st to the 950th realisations. Similarly, the dots are values of the latent process  $W_t$  when it equals  $Y_t$ . As can be seen in the illustration, on the left, when the realisations are high and coming from

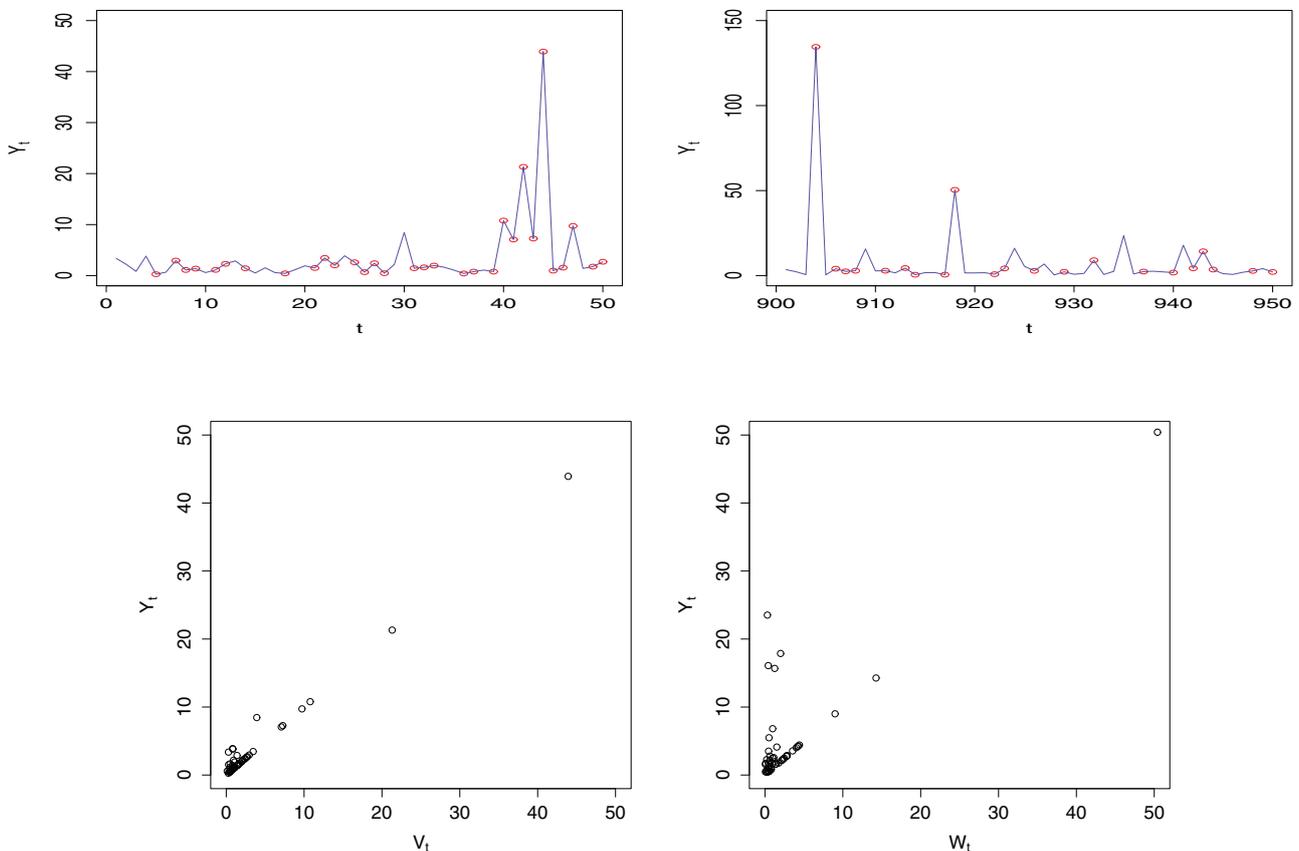


**Figure 2.** Simulation of models (4) (left) and (5) (right) with size = 1000,  $(C, \psi, \beta, \tau) = (0.001, 2, 1.5, -0.5)$  and  $\alpha = (0.5, 0.3, 0.2)$ . The horizontal lines are the 90th percentiles.

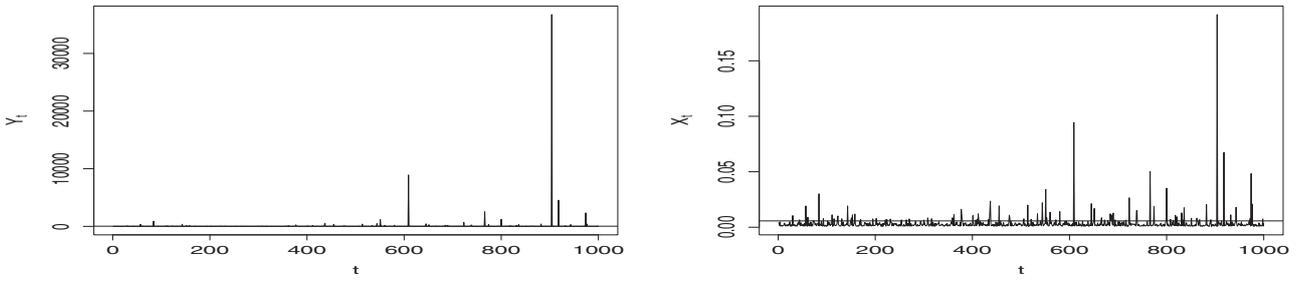
$V_t$ , they are high for extended periods. Although this will not always be the case depending on the values of scaling parameters in  $\alpha$  and the random effects coming from the pertinent  $U_{it}$ 's. On the right, it can be observed that extreme realisations are isolated when they come from  $W_t$ . Such a phenomenon is as a result of the independence structure of the  $W_t$  process. At the bottom, on the left of the figure, we present scatter plots of  $Y_t$  vs.  $V_t$  from the first 50 realisations. On the right, we present scatter plots of  $Y_t$  vs.  $W_t$  from the 901st to 950th realisations. Points on the diagonal are indications of when  $Y_t$  equals the latent process, while points above the diagonal are from observations where  $Y_t$  is greater than the latent process. Due to the specification of  $\beta = 1.5$  for the Fréchet process  $W_t$ , the unit Fréchet  $V_t$

dominates in the right tail region, again, see Figure A3. As a result,  $Y_t = V_t$  more often, therefore, there are fewer values above the diagonal on the left than on the right. It should be noted that  $\psi$  changes the impact of extreme realisations in  $Y_t$  significantly in  $X_t$  as can be seen in Figures 2 and 4.

In Figures 4 and 5, we give more illustrations with a change from  $(\beta, \tau) = (1.5, -0.5)$  to  $(\beta, \tau) = (0.7, 0.5)$ .  $V_t$  remains the same as in the previous simulation while  $W_t$  is different due to changes in shape and location parameters although the underlying unit Fréchet process remains the same. Comparing Figures 2 and 4, it can be seen that the values of  $Y_t$  and  $X_t$  are significantly higher in Figure 4. This observation is due to the values of the  $\beta$  and  $\tau$ , but more importantly



**Figure 3.** Simulation of model (4) with size = 1000,  $(\beta, \tau) = (1.5, -0.5)$  and  $\alpha = (0.5, 0.3, 0.2)$ . Left side: First 50 realisations,  $t = 1 : 50$ . Right side: 901st to 950th realisations,  $t = 901 : 950$ .



**Figure 4.** Simulation of models (4) (left) and (5) (right) with size = 1000,  $(C, \psi, \beta, \tau) = (0.001, 2, 0.7, 0.5)$  and  $\alpha = (0.5, 0.3, 0.2)$ . The horizontal lines are the 90th percentiles.

$\beta$  causing a heavier right tail. Comparing Figures 3 and 5, our observations in Figure 3 is quite different from the observations in Figure 5. In Figure 5, with identical time points as in Figure 3, it can be seen in the illustration on the top-left that there are more isolated extreme realisations. This is due to the independent process  $W_t$  dominating in the right tail. On the right, it can still be observed that extreme realisations are isolated when they come from  $W_t$ . At the bottom, since  $Y_t = W_t$  more often, there are more values above the diagonal on the left than on the right unlike in Figure 3.

**2.3. Theoretical properties**

To look at the theoretical properties of our model, we first look at the distribution functions. The marginal

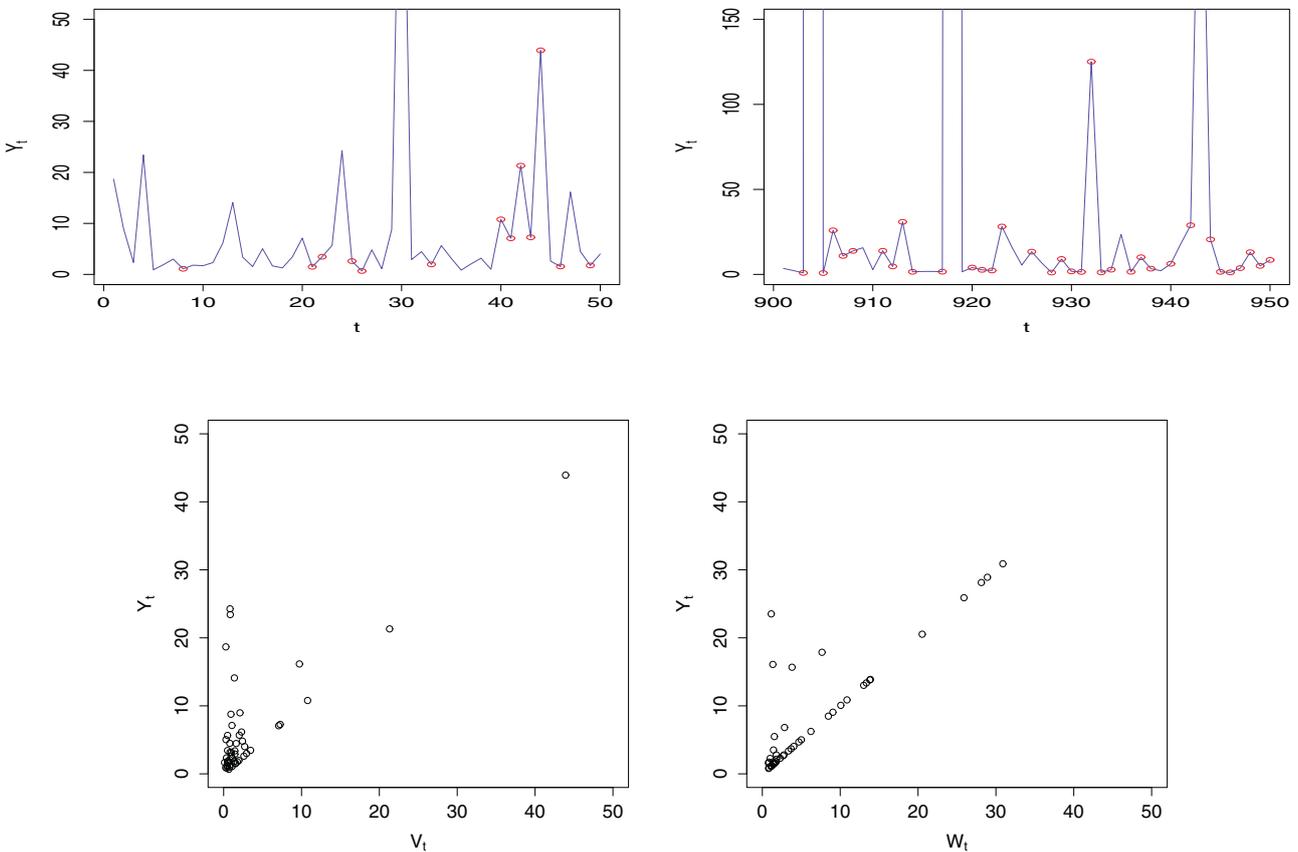
distribution of our model (4) is  $\exp(-\frac{1}{(y_t - \tau)^\beta} - \frac{1}{y_t})$  for any  $y_t > 0$  and any index  $t$ , as shown in Section A.2.1.

The joint distribution takes the form:

$$P(Y_t \leq y_t, 1 \leq t \leq r) = \mathbb{E}[P(Y_t \leq y_t, 1 \leq t \leq r | \mathbf{U})]$$

for  $r \geq 1$  and  $y_t > 0$ .  $\mathbf{U}$  is the set of all relevant random variables on the interval  $[0, 1]$  and

$$P(Y_t \leq y_t, 1 \leq t \leq r | \mathbf{U}) = \exp \left\{ -\frac{r}{(y_t - \tau)^\beta} - \sum_{l=1}^L \alpha_l \left[ \sum_{j=\max(r-l, 0)+1}^r \left( \frac{U_{lj}}{y_j} \right) + \sum_{j=1}^{\max(r-l, 0)} \max \left( \frac{U_{lj}}{y_j}, \frac{(1 - U_{l, j+1})}{y_j^{j+1}} \right) + \sum_{j'=1}^{\min(l, r)} \frac{(1 - U_{lj'})}{y_j^{j'}} \right] \right\}. \tag{6}$$



**Figure 5.** Simulation of model (4) with size = 1000,  $(\beta, \tau) = (0.7, 0.5)$  and  $\alpha = (0.5, 0.3, 0.2)$ . Left side: First 50 realisations,  $t = 1 : 50$ . Right side: 901st to 950th realisations,  $t = 901 : 950$ .  $(Y_{30}, Y_{904}, Y_{918}, Y_{943}) = (110.2, 36749.2, 4545.1, 320.9)$  are out of the y-axis limits.

We note that the marginal does not depend on  $\alpha$ , but the joint does. Hence, when  $\beta < 1$ , implying  $W_t$  is more dominant in the right tail, the estimation of  $\alpha$  becomes harder than the estimation of  $\alpha$  when  $\beta > 1$ , and the difficulty levels increase when  $\beta < 1$  gets smaller. See Section A.2.2 for more detailed steps of the joint distribution.

To investigate the temporal dependence, we apply the asymptotic dependence index among other measures.

**Definition 2.2:** Two random variables  $X_1$  and  $X_2$  with identical distribution function  $F$  are said to be asymptotically dependent (tail dependent) if,

$$\lim_{u \rightarrow x_+} P(X_1 > u | X_2 > u) > 0, \quad (7)$$

where  $x_+ = \sup\{x \in \mathbb{R} : F(x) < 1\}$ . If the limit of the above conditional probability is equal to 0, we say that  $X_1$  and  $X_2$  are asymptotically independent (tail independent).

The above definition of asymptotic dependence between two random variables is due to Sibuya (1959) and an extension given by Ledford and Tawn (2003), Zhang (2005) and Zhang and Huang (2006) is the lag- $r$  asymptotic dependence index defined as follows.

**Definition 2.3:** Let  $X_1, X_2, \dots$  be a stationary time series with distribution function  $F$ . If

$$\lambda_r = \lim_{u \rightarrow x_+} P(X_{r+1} > u | X_1 > u) > 0$$

while

$$\lim_{u \rightarrow x_+} P(X_{r+k} > u | X_1 > u) = 0, \quad k > 1,$$

where  $x_+ = \sup\{x \in \mathbb{R} : F(x) < 1\}$ , we say that the series is asymptotically dependent up to lag- $r$  and  $\lambda_r$  is the lag- $r$  asymptotic dependence index.

Next, we look at the asymptotic properties based on Definition 2.3 above. It should be noted that the asymptotic dependence index of model (4) is identical to that of its counterpart, model (5), as the transformation is continuous.

**Proposition 2.1:** The process (4) is lag- $L$  asymptotically dependent.

**Proposition 2.2:** The process (4) has asymptotic dependence index  $\lambda_l$  as

- $\mathbb{E}[\alpha_l (\min[U_{lt}, (1 - U_{l,t+i})])] \text{ when } \beta > 1;$
- $\frac{\mathbb{E}[\alpha_l (\min[U_{lt}, (1 - U_{l,t+i})])]}{2} \text{ when } \beta = 1;$
- 0 when  $\beta < 1,$

for  $1 \leq l \leq L$ . When  $l > L$ ,  $\lambda_l = 0$ .

The proofs of Propositions 2.1 and 2.2 are provided in Section A.2.4.

### 2.3.1. Stationarity and max-stability

The next properties we look at are the stationarity of model (4) and the max-stability of the sparse M3 part. This is necessary as we aim to model stationary time series data and we need max-stability following the earlier stated results of Smith and Weissman (1996). From (6) it can be seen that the process is stationary as

$$\begin{aligned} P(Y_1 \leq y_1, \dots, Y_r \leq y_r) \\ = P(Y_k \leq y_1, \dots, Y_{k+r-1} \leq y_r) \end{aligned}$$

for any  $r \geq 1$  and  $k \geq 1$ . Also from (6), excluding the random shock part (or when  $\beta = 1$  and  $\tau = 0$ ), it can be seen that the sparse M3 process is max-stable as Definition 2.1 is satisfied.

**Result 2.1:** For any  $x > 0$ , the following holds for model (5),

$$\begin{aligned} P(X_{t+i} > x | X_t > x) \\ = 2 - \frac{1 - \exp\left(-\frac{2}{(y-\tau)^\beta} - \frac{2}{y}\right) 2 \frac{y^2}{\alpha_t^2} \left[\exp\left(\frac{\alpha_t}{y}\right) - \frac{\alpha_t}{y} - 1\right]}{1 - \exp\left(-\frac{1}{(y-\tau)^\beta} - \frac{1}{y}\right)}, \end{aligned}$$

where  $y = \left[\frac{x}{C}\right]^\psi$ .

Result 2.1 can be used to study serial time dependence and we also apply it in our estimation approach for the  $\alpha_t$ 's. The proof of Result 2.1 is in Section A.2.3.

**Result 2.2:** A property of model (5) is that when  $\tau = 0$ ,  $P(X_t \leq C) = \exp(-2)$ .

This is from the following and setting  $\tau = 0$  and  $x_t = C$ .

$$\begin{aligned} P(X_t \leq x_t) &= P\left(Y_t \leq \left[\frac{x_t}{C}\right]^\psi\right) \\ &= \exp\left(-\left(\left[\frac{x_t}{C}\right]^\psi - \tau\right)^{-\beta} - \left[\frac{x_t}{C}\right]^{-\psi}\right). \end{aligned}$$

### 2.4. Estimation approach

There are previous works in the literature on the estimation of parameters of moving maxima processes. For example, Tang et al. (2013) and Zhang and Zhu (2016) used GMM due to Hansen (1982). Bootstrap and empirical distribution techniques have also been proposed by Hall, Peng, and Yao (2002), while Zhang and Smith (2010) proposed the use of signature patterns among other methods. Here, we use Bayesian inference like Chamú-Morales (2005) and Kuniham, Omori, and Zhang (2011). In this section, we discuss relevant Markov chain Monte Carlo (MCMC) methods based on Robert and Casella (2004) and Gelman, Carlin, Stern, and Rubin (2003), and then proceed to our approach.

### Notation

Let  $\Theta^{[S]} = (\Theta^1, \Theta^2, \dots, \Theta^S)$  be a  $d$ -dimensional chain and  $\theta_{1:d}^s = (\theta_1^s, \dots, \theta_d^s)$  be a  $d$ -dimensional vector realisation of  $\Theta^s$  for  $s = 1, \dots, S$ .

Our main focus here is on numerically evaluating integrals of the form

$$\int g(\theta_1, \dots, \theta_d) p(\theta_1, \dots, \theta_d) d\theta_1 \dots d\theta_d, \quad (8)$$

without having to simulate samples directly from  $p$  mostly due to difficulty of doing so.

#### 2.4.1. Hybrid MCMC

Since we have a multidimensional parameter space, we apply the Metropolis–Hasting algorithm from Metropolis et al. (1953) and Hastings (1970) in tandem with the Gibbs sampling technique from Geman and Geman (1984) in what is called the hybrid MCMC shown below in Algorithm 1. Detailed discussion of our estimation approach is given in Appendix A.1.

---

#### Algorithm 1: Hybrid MCMC

For  $k = 1, 2, \dots, d$ , from the most recent value  $(\theta_1^{s+1}, \dots, \theta_{k-1}^{s+1}, \theta_k^s, \dots, \theta_d^s)$ ;

Step1. Generate  $\tilde{\Theta}_k \sim q_k(\theta_k | \theta_1^{s+1}, \dots, \theta_{k-1}^{s+1}, \theta_k^s, \dots, \theta_d^s)$ ;

Step2. Set  $\Theta_k^{s+1}$  as follow:

$$\Theta_k^{s+1} = \begin{cases} \tilde{\Theta}_k, & \text{with probability } P(\theta_k^s, \tilde{\Theta}_k), \\ \theta_k^s, & \text{with probability } 1 - P(\theta_k^s, \tilde{\Theta}_k), \end{cases}$$

where

$$P(\theta_k, \tilde{\theta}_k) = \min \left\{ 1, \frac{p_k(\tilde{\theta}_k | \theta_{1:(k-1)}^{s+1}, \theta_{(k+1):d}^s)}{p_k(\theta_k | \theta_{1:(k-1)}^{s+1}, \theta_{(k+1):d}^s)} \times \frac{q_k(\theta_k | \theta_{1:(k-1)}^{s+1}, \tilde{\theta}_k, \theta_{(k+1):d}^s)}{q_k(\tilde{\theta}_k | \theta_{1:(k-1)}^{s+1}, \theta_k, \theta_{(k+1):d}^s)} \right\}.$$


---

### 3. Simulation

The aim is to see how comparable estimated values are with true values. We have done extensive simulations with various parameter values. In this section, we present results for parameter values close to the estimated values from real market data in Section 4. We simulated a sample from model (5) with a size of  $T = 5,000$  and true parameter values as:

$$\alpha = (\alpha_1, \alpha_2, \alpha_3) = (0.5, 0.3, 0.2),$$

$$C = 0.001, \psi = 2.0, \beta = 1.5, \tau = -0.5.$$

We set  $\delta = 0.2$  for estimation of parameters in  $\alpha$  and the standard deviation for proposals chosen by trial-and-error as  $\sigma_\kappa = 0.02$ ,  $\sigma_\eta = 0.03$ ,  $k_\zeta = 4$ ,  $k'_\zeta = 0.01$ ,  $k_\tau = 5$  and  $k'_\tau = 0.005$ . For  $\rho_l(x^*)$ , we chose  $x^*$  as the 90th percentile. The values were chosen so that we can

have an acceptance rate of about 40% each time we attempted to change a transformation parameter and about 25% for  $\alpha$  (two parameters) as recommended by Gelman, Roberts, and Gilks (1996), Gelman et al. (2003) for multidimensional problems. We ran the chain for 210,000 iterations and then discarded the first 10,000 as burn-in period. Due to autocorrelation, we thinned the remaining 200,000 by picking every 100th realisations and keeping them as samples drawn.

To monitor convergence, we apply the following diagnostic methods: (1) time plots to show the mixing of the chains; (2) density plots to show the estimated posterior distributions; (3) autocorrelation plots to show the autocorrelation function of the kept samples; and (4) acceptance rate from the original samples, which gives the ratio of the number of times that proposals are accepted to the number of iterations.

The MCMC outputs are presented in Figure 6. They include time plots, estimated posterior distributions, and autocorrelation plots. Looking at the outputs and results in Table 1 and Figure 6, we can conclude that our estimation approach works well as we are able to successfully recover true parameter values with reasonable convergence measures.

In our tables,  $\theta_p$  means  $100 \times p$ th percentile and ‘Acc. Rate’ means acceptance rate.

To further test convergence property of our estimation approach, we use the test suggested by Gelman et al. (2003). For  $M > 1$  parallel draws, each of length  $S$ , we label each simulation drawn as  $\{\theta_{sm}, s = 1, \dots, S, m = 1, \dots, M\}$ . Define

$$B = \frac{S}{M-1} \sum_{m=1}^M (\bar{\theta}_{.m} - \bar{\theta}_{..})^2, \quad \text{where,}$$

$$\bar{\theta}_{.m} = \frac{1}{S} \sum_{s=1}^S \theta_{sm}, \quad \bar{\theta}_{..} = \frac{1}{M} \sum_{m=1}^M \bar{\theta}_{.m},$$

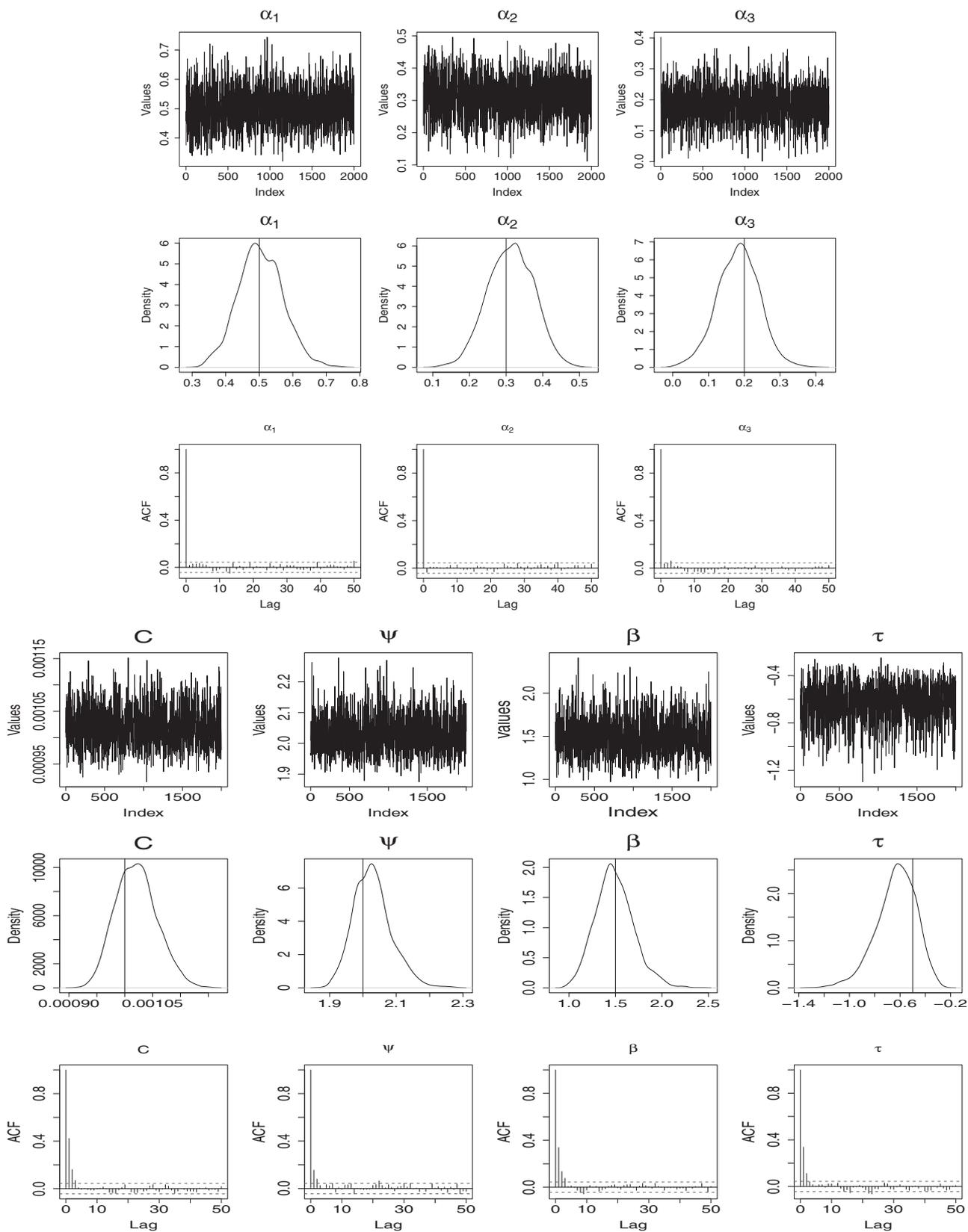
$$W = \frac{1}{M} \sum_{m=1}^M s_m^2, \quad \text{where,}$$

$$s_m^2 = \frac{1}{S-1} \sum_{s=1}^S (\theta_{sm} - \bar{\theta}_{.m})^2,$$

then the variance between ( $B$ ) and within ( $W$ ) sequences are compared. We can estimate the marginal posterior variance,  $V_\theta$  of each parameter, defined as a weighted average of  $B$  and  $W$  in the following form,

$$V_\theta = \frac{S-1}{S} W + \frac{1}{S} B.$$

Convergence can be monitored by checking the factor by which the scale of the current distribution for  $\theta$  can be reduced when the chain is allowed to run indefinitely.



**Figure 6.** Simulation Result 2.1: Graph of MCMC results for all parameters of model (5), showing time plot, estimated posterior distribution and autocorrelation.

The scale factor takes the form,

$$\hat{R} = \sqrt{\frac{V_\theta}{W}}$$

with  $\hat{R} \rightarrow 1$  as the size of the chains goes to  $\infty$ . In Gelman et al. (2003), it was stated that  $\hat{R} < 1.1$  is acceptable for most examples.

Following this approach, we tested convergence as follows. We ran  $M = 10$  chains in parallel using the

**Table 1.** Results of estimation for simulation example of model (5).

$\theta$	True	Mean	Std. Dev.	$\theta_{0.5}$	$[\theta_{0.025}, \theta_{0.975}]$	Acc. Rate
$C$	0.001	0.00102	3.6e-05	0.00102	[0.00095, 0.00109]	0.3972
$\psi$	2.0	2.0289	0.0576	2.0245	[1.9285, 2.1555]	0.3984
$\beta$	1.5	1.4900	0.2067	1.4745	[1.1178, 1.9363]	0.3747
$\tau$	-0.5	-0.6314	0.1570	-0.6188	[-0.9873, -0.3626]	0.3719
$\alpha_1$	0.5	0.5043	0.0660	0.5012	[0.3764, 0.6377]	0.2664
$\alpha_2$	0.3	0.3125	0.0619	0.3138	[0.1927, 0.4298]	0.2664
$\alpha_3$	0.2	0.1831	0.0581	0.1850	[0.0654, 0.2967]	0.2664

same simulated data and set-up as above. We let randomly chosen starting points be dispersed over a certain interval for each parameter. For  $C$ , we started between 0.0008 and 0.0012, and chose a starting point between 1 and 3 for  $\beta$  and  $\psi$ , while we started  $\tau$  between  $-1$  and 0. We started  $\alpha_1$  and  $\alpha_2$  between 0 and 0.5. We let each chain run for 210,000 iterations and discarded the first 10,000 as in the previous example. We, however, did not thin the draws. The  $\hat{R}$  value for each parameter is presented in Table 2. The results suggest that at 200,000 iterations, we attain convergence with each chain.

To check for sensitivity to prior, instead of using the unnormalised uniform priors on  $\kappa, \eta, \zeta$  and  $\tau$  (or logarithmic priors on  $\kappa, \eta$ , and  $\zeta$ ) as used previously, we use uniform priors on  $C, \psi$  and  $\beta$  over the range  $[0, \infty)$ . This implies that the prior densities then take the form  $f_\kappa \propto \exp(\kappa), f_\eta \propto \exp(\eta), f_\zeta \propto \exp(\zeta)$  and for  $\tau$  we leave it as before since it is not reparameterised. Table 3 gives results from estimation on the same data using the newly chosen priors (uniform priors) versus results from the original priors (logarithmic priors). Although both choices of priors are locally uniform, they are different. The logarithmic prior places lower likelihood on larger values, hence penalising larger values, while the uniform prior does not. Based on results in Table 3 and several other runs, we conclude that results from both are comparable.

Finally, as outputs in Table 1 and Figure 6 are based on a single simulated sample, we further examined the performance of our proposed estimation approach for model (5) on multiple samples. The results presented in Table 4 are for 500 different simulated samples each with a size of  $T = 5,000$  and true parameter values chosen as before.

To achieve this, we used parallel computing on a computer with multiple cores. For each sample, since the convergence was quickly reached, we let the chains run for 20,000 iterations and discarded the first 5000.

**Table 2.** Results of convergence test for simulation example of model (5).

$\theta$	$C$	$\psi$	$\beta$	$\tau$	$\alpha_1$	$\alpha_2$	$\alpha_3$
$\hat{R}$	1.00067	1.00036	1.00067	1.00052	1.00006	1.00002	1.00005

**Table 3.** Results of sensitivity analysis for simulation example of model (5).

$\theta$	True	Mean		Median	
		Logarithmic	Uniform	Logarithmic	Uniform
$C$	0.0010	0.00102	0.00102	0.00102	0.00102
$\psi$	2.0	2.0289	2.0286	2.0245	2.0232
$\beta$	1.5	1.4900	1.5158	1.4745	1.5133
$\tau$	-0.5	-0.6314	-0.6516	-0.6188	-0.6357
$\alpha_1$	0.5	0.5043	0.5035	0.5012	0.5021
$\alpha_2$	0.3	0.3125	0.3133	0.3138	0.3139
$\alpha_3$	0.2	0.1831	0.1830	0.1850	0.1812

We thinned the remaining 15,000 iterations by picking every 50th realisations and keeping the observed median values from each sample. Hence, we had a total of 500 estimated medians from each sample. Results of the medians are presented in Table 4. The results further suggest that our estimation approach works well for multiple samples.

For more tests of our estimation approach on model (5), we further tested the performance with the case where  $\beta < 1$ . We simulated a sample from model (5) with a size of  $T = 5000$  and true parameter values as:

$$\alpha = (\alpha_1, \alpha_2, \alpha_3) = (1/3, 1/3, 1/3), C = 0.0012, \psi = 1.5, \beta = 0.7, \tau = 0.5.$$

For estimation, we set  $\delta = 0.6$  for  $\alpha, \sigma_\kappa = 0.02, \sigma_\eta = 0.045, k_\zeta = 10, k'_\zeta = 0.01, k_\tau = 20$  and  $k'_\tau = 0.005$ . For  $\rho_l(x^*)$ , we chose  $x^*$  as the 90th percentile. We ran the chains for 210,000 iterations and then discarded the first 10,000 as burn-in period. Due to autocorrelation, we thinned the remaining 200,000 by picking every 100th realisations and keeping them as samples drawn. Looking at the outputs and results in Table 5 and Figure A4, we can conclude that our estimation approach works well as we have been able to successfully recover true parameter values with reasonable convergence measures.

**Table 4.** Results of estimation for 500 simulation examples, using median values of the parameters from each simulation. [Int. C.P. means Interval Coverage Probability.]

$\theta$	True	Mean	Std. Dev.	$[\theta_{0.025}, \theta_{0.975}]$	95% Int. C.P.
$C$	0.0010	0.00102	4.4e-05	[0.00093, 0.00111]	0.964
$\psi$	2.0	2.0618	0.0759	[1.9393, 2.2316]	0.950
$\beta$	1.5	1.5263	0.2931	[1.0163, 2.1141]	0.939
$\tau$	-0.5	-0.6450	0.1833	[-1.0707, -0.3086]	0.979
$\alpha_1$	0.5	0.5025	0.0615	[0.3833, 0.6262]	0.975
$\alpha_2$	0.3	0.2947	0.0533	[0.1910, 0.3947]	0.981
$\alpha_3$	0.2	0.2026	0.0555	[0.0959, 0.3159]	0.960

#### 4. Real data example

In this section, we carry out analysis of real data. We consider the univariate time series of intra-daily maxima of exchange returns from the US Dollar (USD) versus the Japanese Yen (JPY). The raw data contains 3-minute intra-daily prices for a time period between 31 December 2007, to 27 June 2013. The currencies traded six days weekly from Sunday to Friday. Hence, we have a total of 1720 trading days. There are four missing days that were filled by interpolation. The data was obtained from the University of Wisconsin-Madison Business School database. For each 3-minute period, we derived the negative log-return<sup>2</sup> of the closing price versus the opening price, that is,

$$\text{negative log-return} = -\log \frac{\text{closing price}}{\text{opening price}}.$$

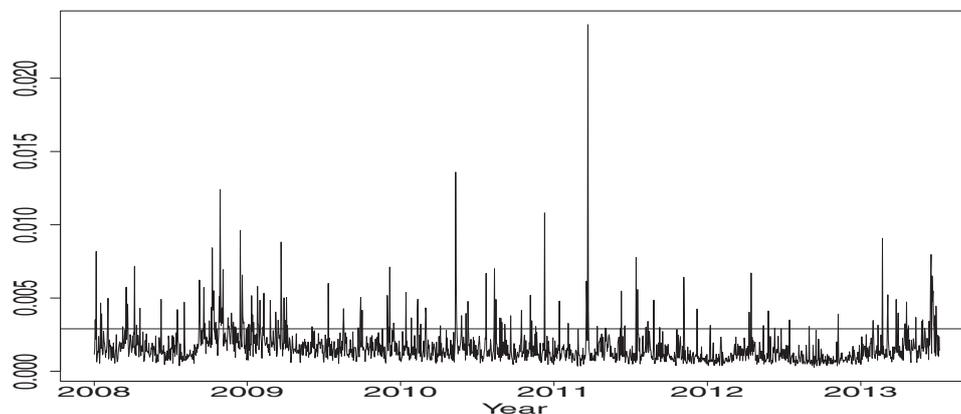
We further derived the daily (block) maxima of these values and used these as the values to which we applied

model (5). By taking block maxima, the distribution of the observations is of the extreme value type. This means we can apply extreme value approaches to model tail dependence, clustering of extremes and to explain the volatility of the series. The time plot of the intra-daily maxima (which looks similar to the plot in Figure 2) is available in Figure 7. We carried out augmented Dickey-Fuller and Phillips-Perron tests, both having  $p$ -values less than 0.01, suggesting that the series is stationary as it appears to be. For the lag value  $L$ , we chose  $L = 6$  as there are six trading days each week. Hence, we can capture up to weekly lag effects.

We apply our estimation approach to estimate parameters that best fit the data. The values  $\delta = 0.4$ ,  $\sigma_\kappa = 0.035$ ,  $\sigma_\eta = 0.05$ ,  $k_\zeta = 2$ ,  $k'_\zeta = 0.01$ ,  $k_\tau = 3$  and  $k'_\tau = 0.005$  were chosen so that we could have an acceptance rate of about 40% each time we attempted to change a transformation parameter and about 25% for  $\alpha$  (5 parameters) as recommended by Gelman et al. (1996) and Gelman et al. (2003) for multidimensional

**Table 5.** Results of estimation for simulation example of model (5).

$\theta$	True	Mean	Std. Dev.	$\theta_{0.5}$	$[\theta_{0.025}, \theta_{0.975}]$	Acc. Rate
$C$	0.0012	0.00122	5.9e-05	0.00123	[0.00107, 0.00131]	0.3761
$\psi$	1.5	1.6163	0.1951	1.6444	[1.1851, 1.9419]	0.4148
$\beta$	0.7	0.6766	0.1214	0.6436	[0.5261, 1.0192]	0.3748
$\tau$	0.5	0.4647	0.0519	0.4591	[0.3757, 0.5740]	0.3635
$\alpha_1$	1/3	0.1851	0.1285	0.1623	[0.0092, 0.4852]	0.4685
$\alpha_2$	1/3	0.3921	0.1712	0.3910	[0.0710, 0.7359]	0.4685
$\alpha_3$	1/3	0.4226	0.1725	0.4202	[0.0786, 0.7686]	0.4685



**Figure 7.** Intra-daily maxima of the US Dollars vs. Japanese Yen 3-minute negative log-returns. The horizontal line is the 90th percentile.

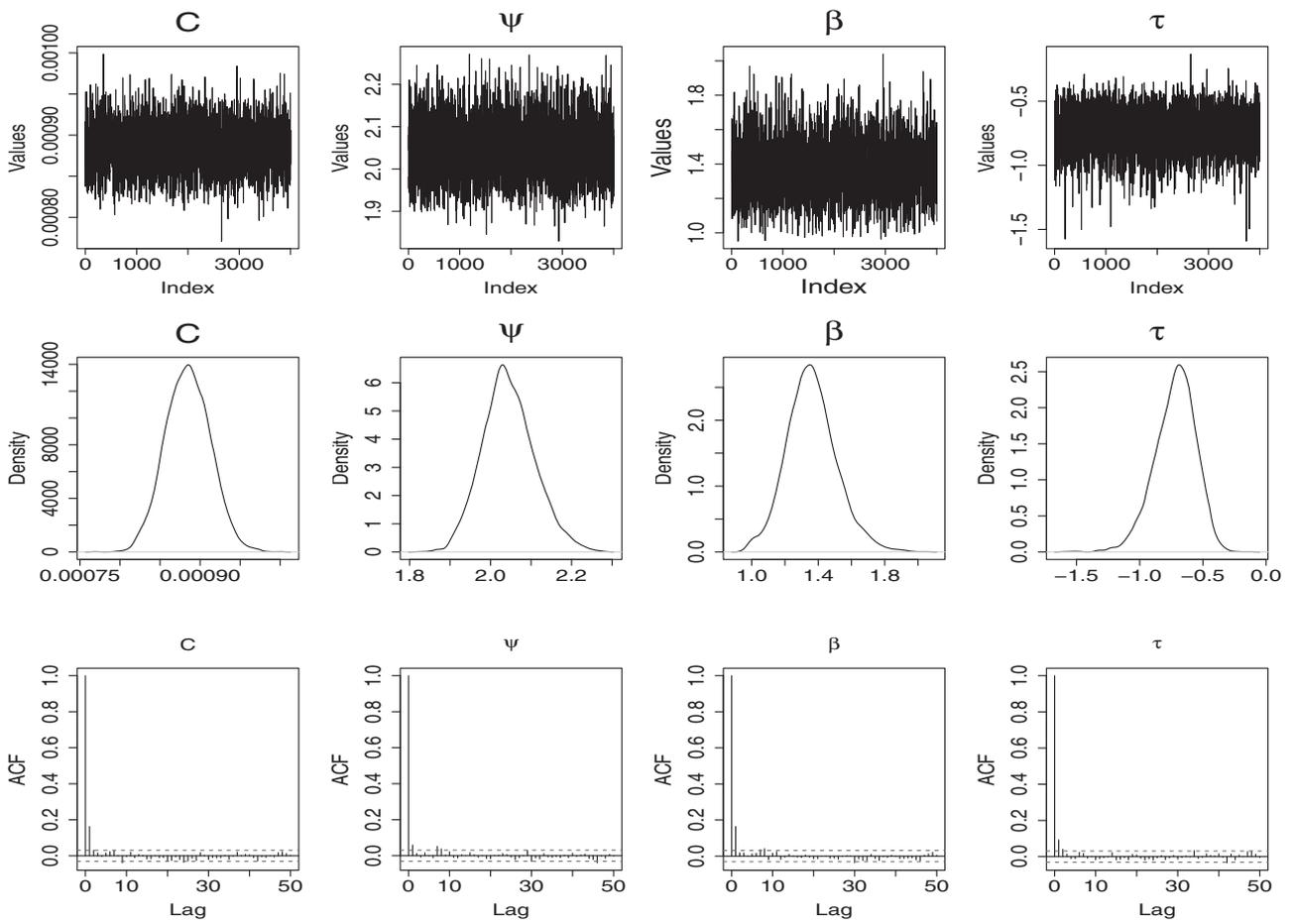


Figure 8. Real data analysis of USD vs JPY: Graph of MCMC results for transformation parameters of model (5), showing time plot, estimated posterior distribution and autocorrelation.

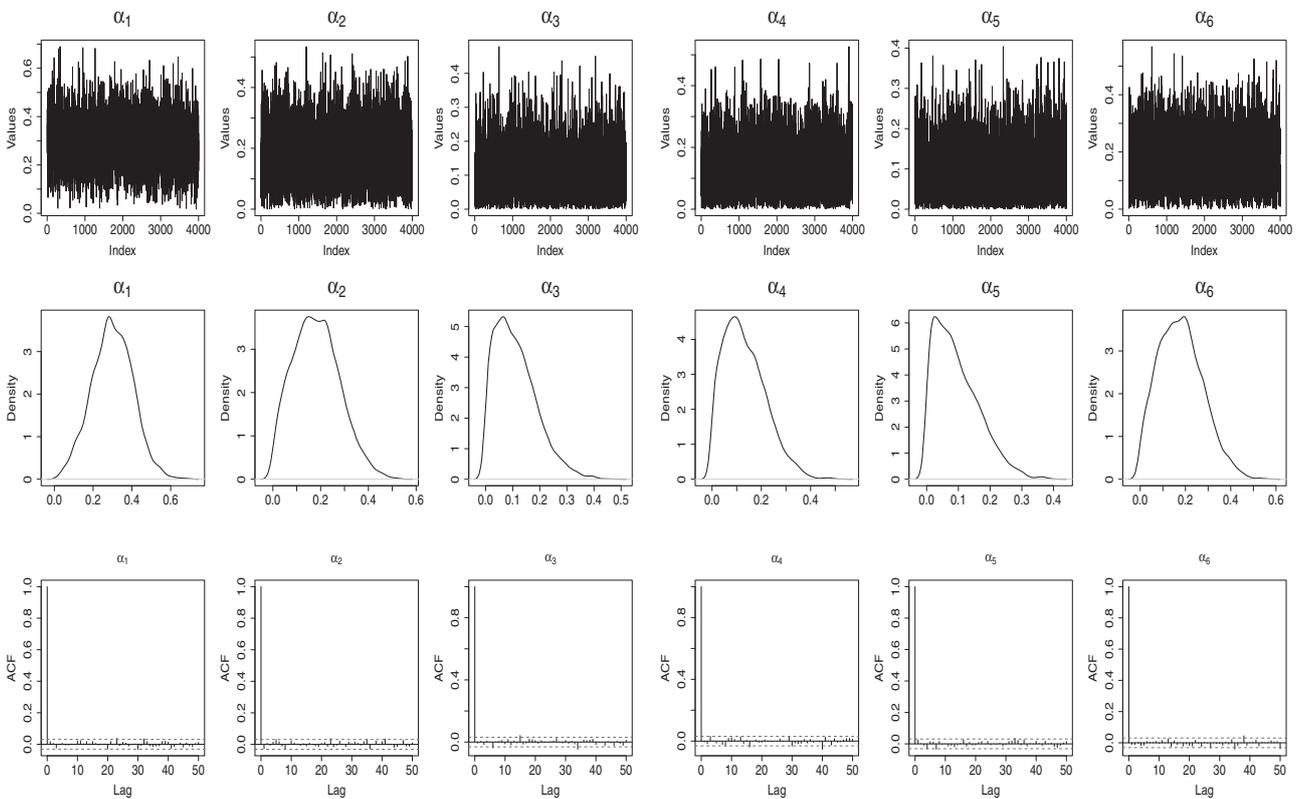


Figure 9. Real data analysis of USD vs JPY: Graph of MCMC results for alpha parameters of model (5), showing time plot, estimated posterior distribution and autocorrelation.

**Table 6.** Results of estimation for real data example of intra-daily maxima of 3-minute negative log-returns of USD vs. JPY.

$\theta$	Mean	Std. Dev.	$\theta_{0.5}$	$[\theta_{0.025}, \theta_{0.975}]$	Acc. Rate
$C$	8.84e-04	2.78e-05	8.84e-04	[8.31e0-4, 9.39e0-4]	0.3959
$\psi$	2.0448	0.0641	2.0406	[1.9264, 2.1825]	0.3926
$\beta$	1.3605	0.1507	1.3534	[1.0768, 1.6860]	0.3595
$\tau$	-0.7200	0.1625	-0.7062	[-1.0650, -0.4444]	0.3805
$\alpha_1$	0.2989	0.1048	0.2973	[0.0938, 0.5059]	0.2533
$\alpha_2$	0.1822	0.0966	0.1778	[0.0180, 0.3861]	0.2533
$\alpha_3$	0.1115	0.0775	0.0996	[0.0052, 0.2938]	0.2533
$\alpha_4$	0.1323	0.0851	0.1200	[0.0077, 0.3267]	0.2533
$\alpha_5$	0.0959	0.0710	0.0818	[0.0039, 0.2623]	0.2533
$\alpha_6$	0.1789	0.0975	0.1732	[0.0138, 0.3879]	0.2533

problems. For  $\rho_l(x^*)$ , we chose  $x^*$  as the 85th percentile. We ran the chain for 210,000 iterations and then discarded the first 10,000 as burn-in period. We thinned the remaining 200,000 by picking every 50th realisations and keeping them as samples drawn.

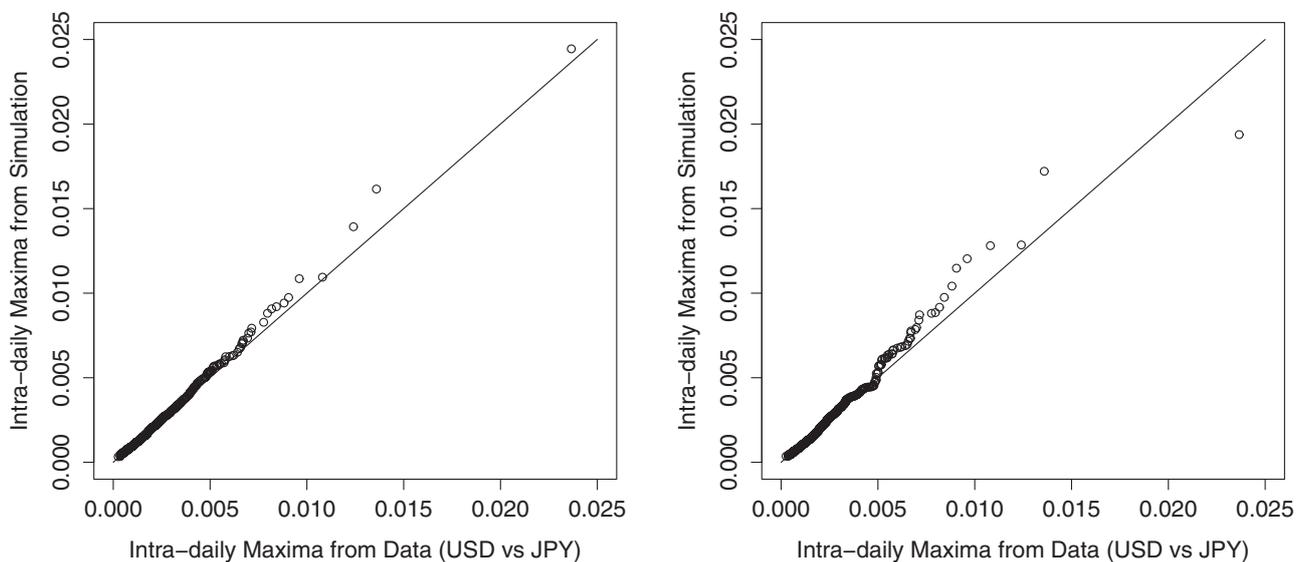
Our results are presented in Table 6. The MCMC outputs are presented in Figures 8 and 9. They include time plots, estimated posterior distributions and autocorrelation plots. Looking at the outputs in Figures 8 and 9, we can conclude that our estimation approach works well with reasonable convergence measures. From the posterior distributions in Table 6, the shape and scale parameters are significantly different from 1 as the credible intervals do not contain 1. Also, the location parameter  $\tau$  is significantly different from 0, as the posterior mean of  $\tau$  is  $-0.72$ , and the credible interval does not contain 0. From probabilistic properties of model 4 and the values of the estimated parameters, the random shock takes effect ( $W_t > V_t$ ) about 30% of the time. To see how our estimated values fit the data, we use a quantile-quantile plot of real data versus simulated data using mean values and median values juxtaposed in Figure 10. The plots suggest that our estimates fit the data well including in the right tail which is the main objective.

## 5. Conclusion

We are able to develop a model for financial time series, starting with the M3 process. Our model (5) can directly fit the data type without any transformations. The model also deals with random signature patterns and has sparsity in its loading matrix to better mirror real data. Given moving maxima models were often applied to filters obtained from some assumed processes, our proposed model is widely applicable.

Our estimation approach has been demonstrated to work. The use of adaptive jumping rules leads to the convergence of Markov chains. We conclude that adaptively scaling jumping kernels by the magnitude of current states gives stability to the samples drawn by the MCMC algorithm. Our work also shows that Bayesian inference can be useful in extreme value theory and estimation in parametric moving maxima processes.

For future work, we could consider similar models applicable to non-stationary series. This task can be achieved for example by having  $C$  to be time dependent, hence as  $C_t$ . Also, we could consider cases where the lag dependence structure in our sparse M3 process is different, and the random coefficients could take other distributions on the  $[0,1]$  interval. For example, based on


**Figure 10.** Quantile-quantile plots using estimated mean values on left and median values on right.

Beta or Dirichlet distributions. Our model can also be extended to the multivariate case building on the multivariate maxima of moving maxima process (M4).

## Notes

1. The process  $X_t = \max\{\alpha Z_t, (1 - \alpha)Z_{t-1}\}$  is called an MM(1) process where  $-\infty < t < \infty$ ,  $\alpha \in (0, 1)$ ,  $Z_t$ 's are unit Fréchet random variables. If  $Z_{t'}$  is considerably larger than  $Z_{t'+1}$  and  $Z_{t'-1}$  then we would have  $\frac{X_{t'+1}}{X_{t'}} = \frac{(1-\alpha)Z_{t'}}{\alpha Z_{t'}} = \frac{(1-\alpha)}{\alpha}$ .
2. A nice property of log-return is its time-additivity. Let  $r_{t,k} = \log\left(\frac{p_t}{p_{t-k}}\right)$ , then  $r_{t,k} = \log\left(\frac{p_t}{p_{t-1}} \frac{p_{t-1}}{p_{t-2}} \dots \frac{p_{t-k+1}}{p_{t-k}}\right) = r_{t,1} + r_{t-1,1} + \dots + r_{t-k,1}$ .
3. <sup>3</sup>The subcase where  $C > \min(X)$  and  $\tau \geq 1$  is not possible under the constraint.

## Acknowledgments

The authors thank the editor and referees for their valuable comments and suggestions. The work is partially supported by NSF-DMS-1505367 and NSF-CMMI-1536978.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

NSF-DMS-1505367; NSF-CMMI-1536978.

## Notes on contributors

**Timothy Idowu** earned a PhD in statistics from the University of Wisconsin-Madison. He currently works in quantitative finance in the Greater New York City area. Timothy was a doctoral student of the co-author Zhengjun Zhang.

**Zhengjun Zhang** earned a PhD in statistics from the University of North Carolina at Chapel Hill. He is currently a full professor of statistics at the University of Wisconsin-Madison where he has taught statistics and supervised several doctoral students for over a decade.

## References

- Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Chamú-Morales, F. (2005). *Estimation of max-stable processes using Monte Carlo methods with applications to financial risk assessment* (PhD dissertation). Department of Statistics at the University of North Carolina, Chapel Hill, NC.
- Davis, R. A., & Resnick, S. I. (1989). Basic properties and prediction of Max-ARMA processes. *Advances in Applied Probability*, 21(4), 781–803.
- Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis*, 13(2), 257–272.
- Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of Cambridge Philosophical Society*, 24 (2), 180–190.
- Gelman, A., Roberts, G. O., & Gilks, W. R. (1996). Efficient metropolis jumping rules. *Bayesian Statistics*, 5, 599–607.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis, 2e*. Boca Raton, FL: CRC Press.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire [On the limiting distribution of the maximum term in a random series]. *Annals of Mathematics*, 44(3), 423–453.
- Hall, P., Peng, L., & Yao, Q. (2002). Moving-maximum models for extrema of time series. *Journal of Statistical Planning Inference*, 103(1–2), 51–63.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimator. *Econometrica*, 50(4), 1029–1054.
- Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their application. *Biometrika*, 57(1), 97–109.
- Heffernan, J. E., Tawn, J. A., & Zhang Z., (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes. *Extremes*, 10(1), 57–82.
- Kunihama, T., Omori, Y., & Zhang Z., (2011). Efficient estimation and particle filter for max-stable processes. *Journal of Time Series Analysis*, 33(1), 61–80.
- Ledford, A. W., & Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2), 521–543.
- McNeil, A. J., & Frey R., (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7(3–4), 271–300.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller E., (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.
- Pasarica, C., & Gelman, A. (2010). Adaptively scaling the metropolis algorithm using expected squared jumped distance. *Statistica Sinica*, 20, 343–364.
- Robert, C. P., & Casella, G. (2004). *Monte carlo statistical methods, 2e*. New York, NY: Springer Text in Statistics.
- Sibuya, M. (1959). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2), 195–210.
- Smith, R. L., & Weissman, I. (1996). *Characterization and estimation of the multivariate extremal index*. (Tech. Report). Chapel Hill, NC: Department of Statistics at the University of North Carolina.
- Tang, R., Shao, J., & Zhang, Z. (2013). Sparse moving maxima models for tail dependence in multivariate financial time series. *Journal of Statistical Planning and Inference*, 143(5), 882–895.
- Zhang, Z. (2005). A new class of tail-dependent time series models and its applications in financial time series. *Advances in Econometrics*, 20(B), 323–358.
- Zhang, Z., & Huang, J. (2006). Extremal financial risk model and portfolio evaluation. *Computational Statistics and Data Analysis*, 51(4), 2313–2338.
- Zhang, Z., & Smith, R. L. (2010). On the estimation and application of max-stable processes. *Journal of Statistical Planning and Inference*, 140, 1135–1153.
- Zhang, Z., & Zhu, B. (2016). Copula structured M4 processes with application to high-frequency financial data. *Journal of Econometrics*, 194, 231–241.

## Appendix

### A.1. Estimation of parameters

Here, we give details of our estimation approach to the directly applicable model (5). The goal is to draw samples from the joint posterior distribution using Algorithm 1.

The target distribution is the joint posterior distribution of parameters given observations  $\mathbf{x} = (x_1, \dots, x_T)$ . The full set of parameters is:

$$\theta = (\alpha, \log C, \log \psi, \log \beta, \tau),$$

hence,

$$p(\theta|\mathbf{x}) = p(\alpha, \log C, \log \psi, \log \beta, \tau|\mathbf{x}),$$

$$\mathbf{x} = (x_1, \dots, x_T). \quad (\text{A.1})$$

Based on Bayes' theorem, we have:

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\int p(\mathbf{x}|\theta)p(\theta)d(\theta)} \propto p(\mathbf{x}|\theta)p(\theta), \quad (\text{A.2})$$

where  $p(\mathbf{x}|\theta)$  is the likelihood of  $\theta$  and  $p(\theta)$  is its prior density with prior independence assumed among the parameters. We begin with some initial value  $\theta^1$  that satisfies model constraints. In our examples, we use  $\alpha_l^1 = 1/L$  for  $l = 1, \dots, L$ ,  $\psi^1 = 1$ ,  $\beta^1 = 1$ ,  $\tau^1 = 0$  and  $C^1 = \exp(-2) \times 100$ th percentile of the data based on Result 2.2. Numerically obtained maximum likelihood estimates can also be used as there are no analytic solutions.

We then iteratively draw  $S$  samples from  $\theta|\mathbf{x}$  with steps as follows and densities later defined:

$$\begin{aligned} \alpha^{s+1} &\sim p_1(\alpha|C^s, \psi^s, \beta^s, \tau^s, \mathbf{x}), \\ C^{s+1} &\sim p_2(C|\alpha^{s+1}, \psi^s, \beta^s, \tau^s, \mathbf{x}), \\ \psi^{s+1} &\sim p_3(\psi|\alpha^{s+1}, C^{s+1}, \beta^s, \tau^s, \mathbf{x}), \\ \beta^{s+1} &\sim p_4(\beta|\alpha^{s+1}, C^{s+1}, \psi^{s+1}, \tau^s, \mathbf{x}), \\ \tau^{s+1} &\sim p_5(\tau|\alpha^{s+1}, C^{s+1}, \psi^{s+1}, \beta^{s+1}, \mathbf{x}), \end{aligned}$$

for  $s = 1, \dots, S$ .

**A.1.1. Sampling  $\alpha$ .** For  $\alpha = (\alpha_1, \dots, \alpha_L)$  with dimension  $L$  (model is up to lag- $L$ ), we assume a uniform prior, that is,  $f(\alpha) \propto \text{constant}$ . For each of the first  $L - 1$  components of  $\alpha$ ,  $(\alpha_1, \dots, \alpha_{L-1})$ , we generate proposals as follows. For state  $s + 1$  with the current state as  $s$ , for each  $l = 1, \dots, L - 1$ , generate  $\tilde{\alpha}_l \sim q_1(\alpha_l|\alpha_l^s)$  where  $q_1(\cdot|\alpha_l^s)$  is the density function of the uniform distribution,  $Unif(\max[\alpha_l^s - \delta, 0], \min[\alpha_l^s + \delta, 1])$ , with  $\delta$  as a fixed disturbance. When we successfully simulate  $(\tilde{\alpha}_1, \dots, \tilde{\alpha}_{L-1})$  such that  $\sum_{l=1}^{L-1} \tilde{\alpha}_l < 1$ , we set  $\tilde{\alpha}_L = 1 - \sum_{l=1}^{L-1} \tilde{\alpha}_l$ . We therefore have

$$q_1(\cdot|\alpha^s) = \prod_{l=1}^{L-1} \frac{1}{\min[\alpha_l^s + \delta, 1] - \max[\alpha_l^s - \delta, 0]},$$

which implies

$$\frac{q_1(\alpha^s|\tilde{\alpha})}{q_1(\tilde{\alpha}|\alpha^s)} = \prod_{l=1}^{L-1} \frac{\min[\alpha_l^s + \delta, 1] - \max[\alpha_l^s - \delta, 0]}{\min[\tilde{\alpha}_l + \delta, 1] - \max[\tilde{\alpha}_l - \delta, 0]}.$$

For the likelihood, we use the tail dependence properties through the conditional probabilities of exceeding a given threshold. For a high threshold value  $x^*$ , we introduce,  $\rho_l(x^*) = \rho_l(x^*; \alpha, C, \psi, \beta, \tau)$ ,  $n_l(x^*)$  and  $\omega_l(x^*)$  defined as follows:

$$\begin{aligned} \rho_l(x^*) &= P(X_{t+l} > x^* | X_t > x^*) \\ &= P\left(Y_{t+l} > \left[\frac{x^*}{C}\right]^\psi \mid Y_t > \left[\frac{x^*}{C}\right]^\psi\right); \\ n_l(x^*) &= \sum_{t=1}^{T-l} I(X_t > x^*) \\ &= \sum_{t=1}^{T-l} I\left(Y_t > \left[\frac{x^*}{C}\right]^\psi\right); \\ \omega_l(x^*) &= \sum_{t=1}^{T-l} I(X_{t+l} > x^* | X_t > x^*) \\ &= \sum_{t=1}^{T-l} I\left(Y_{t+l} > \left[\frac{x^*}{C}\right]^\psi \mid Y_t > \left[\frac{x^*}{C}\right]^\psi\right) \\ &= \sum_{t=1}^{T-l} I_t^l(x^*), \end{aligned}$$

where  $I$  is the indicator function.  $\mathbb{E}[I_t^l(x^*)] = \rho_l(x^*)$  for each  $t = 1, \dots, T - l$ .

Calculation of  $\rho_l(x^*)$  is based on Result 2.1. From many simulations, we find  $x^*$  equal to the 85th, 90th and 95th percentiles appropriate.

With  $\tilde{\rho}_l(x^*) = \rho_l(x^*; \tilde{\alpha}, C, \psi, \beta, \tau)$  and treating the  $I_t^l(x^*)$ 's as Bernoulli random variables with the probability of success as  $\rho_l(x^*)$ , we have the likelihood for this step as:

$$p_1(\mathbf{x}|\tilde{\alpha}, C, \psi, \beta, \tau) = \prod_{l=1}^L \tilde{\rho}_l(x^*)^{\omega_l(x^*)} (1 - \tilde{\rho}_l(x^*))^{[n_l(x^*) - \omega_l(x^*)]}, \quad (\text{A.3})$$

$$p_1(\mathbf{x}|\alpha, C, \psi, \beta, \tau) = \prod_{l=1}^L \rho_l(x^*)^{\omega_l(x^*)} (1 - \rho_l(x^*))^{[n_l(x^*) - \omega_l(x^*)]}. \quad (\text{A.4})$$

We therefore accept  $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_L)$  with probability:

$$P(\alpha^s, \tilde{\alpha}) = \min \left\{ 1, \frac{p_1(\mathbf{x}|\tilde{\alpha}, C, \psi, \beta, \tau) q_1(\alpha^s|\tilde{\alpha})}{p_1(\mathbf{x}|\alpha, C, \psi, \beta, \tau) q_1(\tilde{\alpha}|\alpha^s)} \right\}. \quad (\text{A.5})$$

**A.1.2. Sampling  $(\log C, \log \psi, \log \beta, \tau)$ .** Set  $\kappa = \log C$ ,  $\eta = \log \psi$  and  $\zeta = \log \beta$ . For all the steps, we use unnormalised uniform priors. The prior densities  $f_\kappa, f_\eta,$

$f_\zeta$  and  $f_\tau$  are therefore flat over the range of possible values for each parameter. For the likelihood we have:

$$\begin{aligned} p(\mathbf{x}|\kappa, \eta, \zeta, \tau) &= p(\mathbf{x}|C, \psi, \beta, \tau) \\ &= \prod_{t=1}^T \exp \left\{ - \left( \left[ \frac{x_t}{C} \right]^\psi - \tau \right)^{-\beta} - \left[ \frac{x_t}{C} \right]^{-\psi} \right\} \times \end{aligned} \quad (\text{A.6})$$

$$\left[ \beta \left( \left[ \frac{x_t}{C} \right]^\psi - \tau \right)^{-\beta-1} + \left[ \frac{x_t}{C} \right]^{-2\psi} \right] \left[ \frac{\psi}{C} \left[ \frac{x_t}{C} \right]^{\psi-1} \right],$$

defined on the set  $\{x : [\frac{x}{C}]^\psi - \tau > 0\}$  with  $C > 0$ ,  $\psi > 0$ ,  $\beta > 0$  and  $\tau \in \mathbb{R}$ .

Given the constraint  $\{x : [\frac{x}{C}]^\psi - \tau > 0\}$ , we look at the range of values such that each parameter satisfies the constraint. We define  $\min(X)$  as the smallest value in our observed data. We obtain the following ranges for parameters.

- $C > 0$ : Two cases,  $\tau \leq 0$  and  $\tau > 0$ .  
Case 1: When  $\tau \leq 0$ ;  
Any  $C > 0$  satisfies the constraint  $\Rightarrow \log(C) \in \mathbb{R}$ .  
Case 2: When  $\tau > 0$ ;  
 $C < \frac{\min(X)}{\tau^{1/\psi}} \Rightarrow \log(C) < \log \left[ \frac{\min(X)}{\tau^{1/\psi}} \right]$ .
- $\psi > 0$ : Two cases.  $\tau \leq 0$  and  $\tau > 0$ .  
Case 1: When  $\tau \leq 0$ ;  
Any  $\psi > 0$  satisfies the constraint  $\Rightarrow \log(\psi) \in \mathbb{R}$ .  
Case 2: When  $\tau > 0$ ; Three subcases<sup>3</sup>;  
  
Subcase 1:  $C < \min(X)$  and  $0 < \tau < 1$ ;  
Any  $\psi > 0$  satisfies the constraint  $\Rightarrow \log(\psi) \in \mathbb{R}$ .  
Subcase 2:  $C < \min(X)$  and  $\tau \geq 1$ ;  
 $\psi > \frac{\log(\tau)}{\log(\frac{\min(X)}{C})} \Rightarrow \exp(\eta) > \frac{\log(\tau)}{\log(\frac{\min(X)}{\exp(\kappa)})}$ .  
Subcase 3:  $C > \min(X)$ ;  
 $\psi < \frac{\log(\tau)}{\log(\frac{\min(X)}{C})} \Rightarrow \exp(\eta) < \frac{\log(\tau)}{\log(\frac{\min(X)}{\exp(\kappa)})}$ .
- $\beta > 0$ ;  
Any  $\beta > 0$  satisfies the constraint  $\Rightarrow \log(\beta) \in \mathbb{R}$ .
- $\tau \in \mathbb{R}$ ;  
 $\tau < \left[ \frac{\min(X)}{C} \right]^\psi \Rightarrow \tau < \left[ \frac{\min(X)}{\exp(\kappa)} \right]^{\exp(\eta)}$ .

We define  $\mathcal{TN}(\mu, \sigma^2, a, b)$  as a truncated normal distribution with mean  $\mu$ , variance  $\sigma^2$ , lower bound  $a$  and upper bound  $b$ . For state  $s+1$  with the current state as  $s$ , simulate proposals for each parameter based on data and the current states of parameters as follows:

- (1)  $\tilde{\kappa} \sim q_2(\kappa|\kappa^s)$  which is  $\mathcal{TN}(\kappa^s, \kappa_\sigma^2, -\infty, \log[\frac{\min(X)}{\tau^{1/\psi}}])$  when  $\tau > 0$  and  $\mathcal{N}(\kappa^s, \kappa_\sigma^2)$  when  $\tau \leq 0$ , the standard deviation  $\kappa_\sigma$  is fixed so that the acceptance rate is within the desired interval.

Accept  $\tilde{\kappa}$  with probability:

$$\min \left\{ 1, \frac{p(\mathbf{x}|\tilde{\kappa}, \eta^s, \zeta^s, \tau^s) f_\kappa(\tilde{\kappa}) q_2(\kappa^s|\tilde{\kappa})}{p(\mathbf{x}|\kappa^s, \eta^s, \zeta^s, \tau^s) f_\kappa(\kappa^s) q_2(\tilde{\kappa}|\kappa^s)} \right\}.$$

- (2)  $\tilde{\eta} \sim q_3(\eta|\eta^s)$  which is  $\mathcal{N}(\eta^s, \eta_\sigma^2)$ , the standard deviation  $\eta_\sigma$  is fixed so that the acceptance rate is within the desired interval except in the following cases.

Case 1:  $\exp(\kappa^s) < \min(X)$  and  $\tau^s \geq 1$

$\exp(\tilde{\eta}) \sim q_3(\exp(\eta)|\exp(\eta^s))$  which is  $\mathcal{TN}(\exp(\eta^s), \eta_\sigma^2, \frac{\log(\tau)}{\log(\frac{\min(X)}{\exp(\kappa^s)})}, \infty)$ , the standard deviation  $\eta_\sigma$  is fixed so that the acceptance rate is within the desired interval.

Case 2:  $\exp(\kappa^s) > \min(X)$

$\exp(\tilde{\eta}) \sim q_3(\exp(\eta)|\exp(\eta^s))$  which is  $\mathcal{TN}(\exp(\eta^s), \eta_\sigma^2, 0, \frac{\log(\tau)}{\log(\frac{\min(X)}{\exp(\kappa^s)})})$ , the standard deviation  $\eta_\sigma$  is fixed so that the acceptance rate is within the desired interval.

Accept  $\tilde{\eta}$  (or  $\exp(\tilde{\eta})$ ) with probability:

$$\min \left\{ 1, \frac{p(\mathbf{x}|\kappa^{s+1}, \tilde{\eta}, \zeta^s, \tau^s) f_\eta(\tilde{\eta}) q_3(\eta^s|\tilde{\eta})}{p(\mathbf{x}|\kappa^{s+1}, \eta^s, \zeta^s, \tau^s) f_\eta(\eta^s) q_3(\tilde{\eta}|\eta^s)} \right\}.$$

- (3)  $\tilde{\zeta} \sim q_4(\zeta|\zeta^s)$  which is  $\mathcal{N}(\zeta^s, \zeta_\sigma^2)$ , the standard deviation  $\zeta_\sigma = \frac{|\zeta^s|}{k_\zeta} + k'_\zeta$ .  $k_\zeta > 0$  and  $k'_\zeta > 0$  are fixed so that the acceptance rate is within the desired interval.

Accept  $\tilde{\zeta}$  with probability:

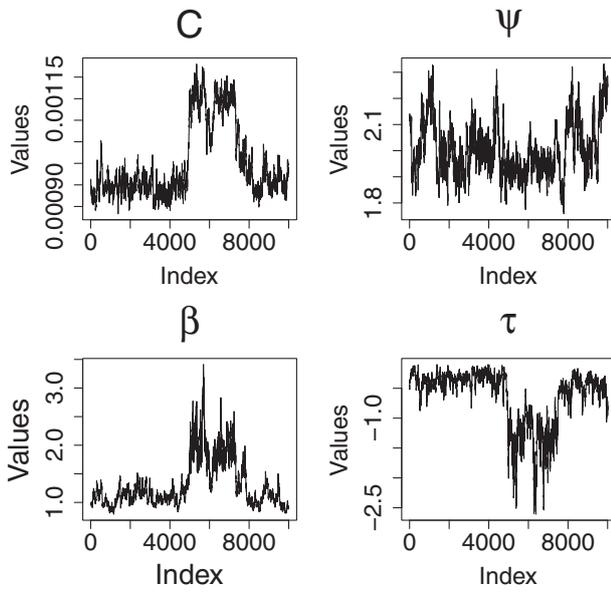
$$\min \left\{ 1, \frac{p(\mathbf{x}|\kappa^{s+1}, \eta^{s+1}, \tilde{\zeta}, \tau^s) f_\zeta(\tilde{\zeta}) q_4(\zeta^s|\tilde{\zeta})}{p(\mathbf{x}|\kappa^{s+1}, \eta^{s+1}, \zeta^s, \tau^s) f_\zeta(\zeta^s) q_4(\tilde{\zeta}|\zeta^s)} \right\}.$$

- (4)  $\tilde{\tau} \sim q_5(\tau|\tau^s)$  which is  $\mathcal{TN}(\tau^s, \tau_\sigma^2, -\infty, \frac{\min(X)}{\exp(\kappa^s)})$ , the standard deviation  $\tau_\sigma = \frac{|\tau^s|}{k_\tau} + k'_\tau$ .  $k_\tau > 0$  and  $k'_\tau > 0$  are fixed so that the acceptance rate is within the desired interval.

Accept  $\tilde{\tau}$  with probability:

$$\min \left\{ 1, \frac{p(\mathbf{x}|\kappa^{s+1}, \eta^{s+1}, \zeta^{s+1}, \tilde{\tau}) f_\tau(\tilde{\tau}) q_5(\tau^s|\tilde{\tau})}{p(\mathbf{x}|\kappa^{s+1}, \eta^{s+1}, \zeta^{s+1}, \tau^s) f_\tau(\tau^s) q_5(\tilde{\tau}|\tau^s)} \right\}.$$

**Use of Adaptive Jumping Kernels.** Adaptively scaling jumping kernels has been shown to improve convergence (see, for example, Pasarica & Gelman, 2010). In items 3 and 4 above, where we sample for  $\zeta$  and  $\tau$ , we use adaptive jumping kernels. The reason for this is the rapidly changing correlations between parameters in the model (for illustration, see Figure A2). The correlations can sometimes lead to rapid increase in the states of  $C$  and  $\beta$  while  $\tau$  decreases rapidly (for illustration, see Figure A1). This is the case even under several re-parameterisations. However, by scaling the jumping kernels by the magnitude of the current state, we preempt such arbitrary fluctuations and



**Figure A1.** Simulation result: Graph of MCMC results showing rapid fluctuation of parameters.

even when they occur, and then the chains quickly move back to more probable states. The role of  $k'_\zeta$  and  $k'_\tau$  is to help the chain move away from states where  $\zeta \approx 0$  ( $\beta \approx 1$ ) and  $\tau \approx 0$ , respectively. Without adaptively scaling jumping kernels, the chain could be stuck within those regions for long periods and even stay still in cases of equality.

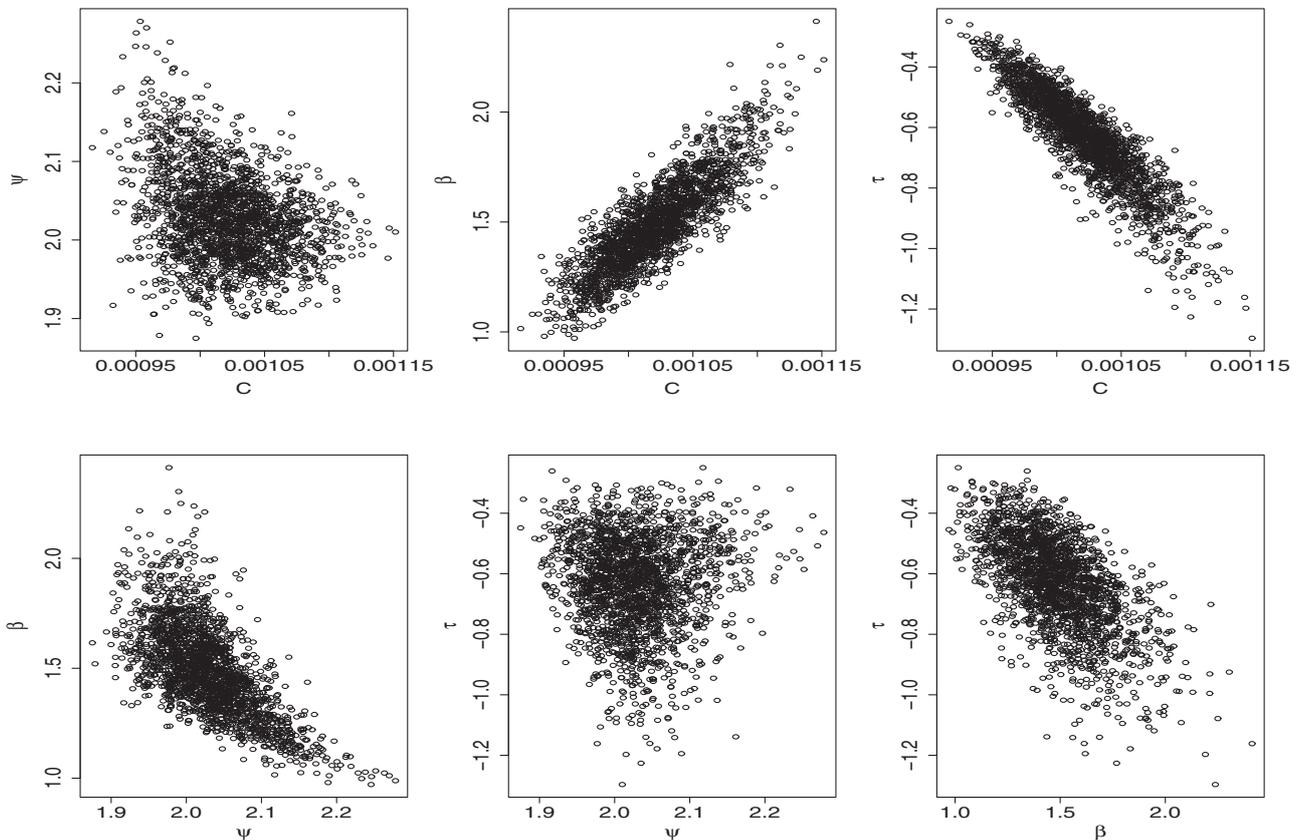
## A.2. Proofs

### A.2.1. Marginal distribution.

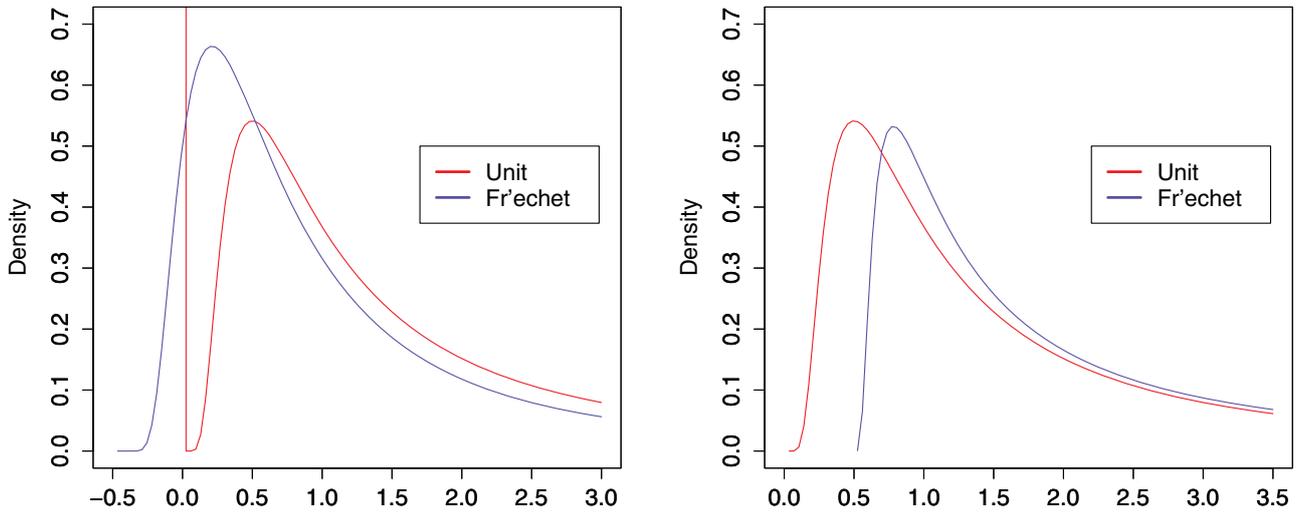
$$\begin{aligned}
 P(X_t \leq x_t) &= P\left(CY_t^{1/\psi} \leq x_t\right) = P\left(Y_t \leq \left[\frac{x_t}{C}\right]^\psi\right) \\
 &= P(Y_t \leq y_t) \\
 &= P(W_t \leq y_t) \cdot P(\max[A_t \cdot Z_t] \leq y_t) \\
 &= \exp\left(-\frac{1}{(y_t - \tau)^\beta}\right) \cdot \prod_{l=1}^L P(\alpha_l U_{lt} Z_{lt} \leq y_t) \\
 &\quad \cdot P(\alpha_l(1 - U_{lt})Z_{l,t-1} \leq y_t) \\
 &= \exp\left(-\frac{1}{(y_t - \tau)^\beta}\right) \\
 &\quad \cdot \mathbb{E}\left(\exp\left(\sum_{l=1}^L \left[-\frac{\alpha_l U_{lt}}{y_t} - \frac{\alpha_l(1 - U_{lt})}{y_t}\right]\right)\right) \\
 &= \exp\left(-\frac{1}{(y_t - \tau)^\beta} - \frac{1}{y_t}\right). \tag{A.7}
 \end{aligned}$$

**A.2.2. Joint distribution.** For  $r \geq 1$ ,  $x_t > 0$  and  $y_t > 0$  where  $t = 1, \dots, r$ .

$$\begin{aligned}
 P(X_t \leq x_t, 1 \leq t \leq r) &= \mathbb{E}\left[P(Y_t \leq y_t, 1 \leq t \leq r | \mathbf{U})\right], \\
 y_t &= \left[\frac{x_t}{C}\right]^\psi,
 \end{aligned}$$



**Figure A2.** Simulation result: Graph of MCMC results showing scatter plot of parameters.



**Figure A3.** Unit Fréchet density (left) and Fréchet( $\beta, 1, \tau$ ) density (right). Left:  $\beta = 1.5$  and  $\tau = -0.5$ . Right:  $\beta = 0.7$  and  $\tau = 0.5$ .

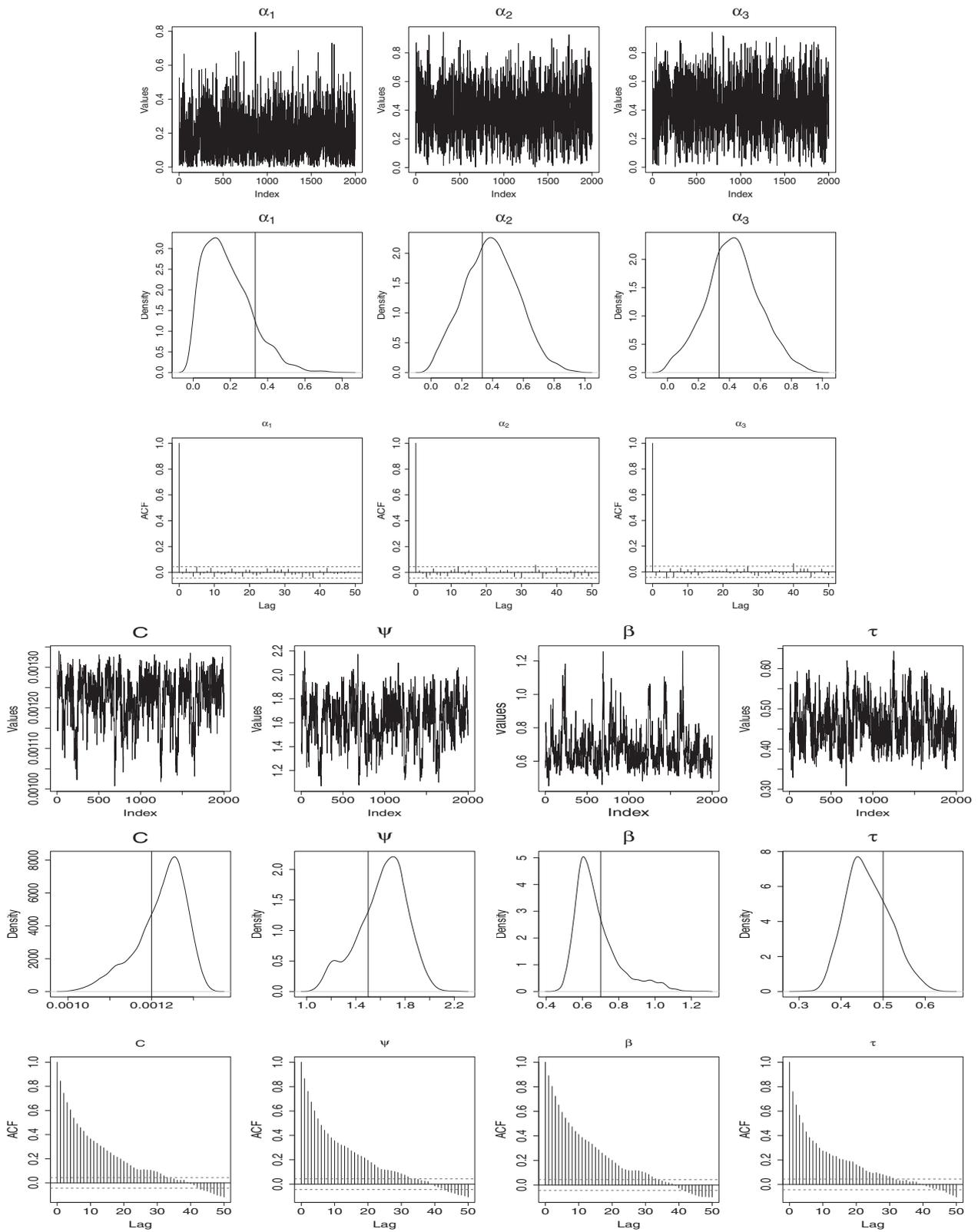
where

$$\begin{aligned}
 P(Y_t \leq y_t, 1 \leq t \leq r | \mathbf{U}) &= P(W_t \leq y_t, 1 \leq t \leq r) \\
 &\cdot P(\max[A_t \cdot Z_t] \leq y_t, 1 \leq t \leq r | \mathbf{U}) \quad (\text{A.8}) \\
 &= \prod_{t=1}^r \exp\left(-\frac{1}{(y_t - \tau)^\beta}\right) \\
 &\times \prod_{l=1}^L \exp\left\{-\left[\sum_{j=\max(r-l, 0)+1}^r \left(\frac{\alpha_l U_{lj}}{y_j}\right)\right.\right. \\
 &+ \sum_{j'=1}^{\max(r-l, 0)} \max\left(\frac{\alpha_l U_{lj'}}{y_{j'}}, \frac{\alpha_l(1 - U_{l, j'+1})}{y_{j'+1}}\right) \\
 &\left.\left. + \sum_{j''=1}^{\min(l, r)} \frac{\alpha_l(1 - U_{lj''})}{y_{j''}}\right]\right\}. \quad (\text{A.9})
 \end{aligned}$$

**A.2.3. Proof of Result 2.1.** For  $1 \leq l \leq L$  and  $y = (x/c)^\psi$ .

$$\begin{aligned}
 \rho_l(x) &= P(X_{t+l} > x | X_t > x) = P(Y_{t+l} > y | Y_t > y). \\
 &= \frac{P(Y_{t+l} > y | Y_t > y)}{P(Y_t > y)} \\
 &= \frac{P(Y_{t+l} > y) + P(Y_t > y) - [1 - P(Y_{t+l} < y, Y_t < y)]}{P(Y_t > y)}. \quad (\text{A.10})
 \end{aligned}$$

$$\begin{aligned}
 P(Y_{t+l} < y, Y_t < y) &= \mathbb{E}[P(Y_{t+l} < y, Y_t < y | \mathbf{U})] \\
 &= \mathbb{E}\left[\exp\left(-\frac{2}{(y - \tau)^\beta}\right)\right. \\
 &\quad \left.- \left[\frac{2(1 - \alpha_l) + \alpha_l((U_{l, t+l} + (1 - U_{lt})) + \max[U_{lt}, (1 - U_{l, t+l}]))}{y}\right]\right)] \\
 &= \mathbb{E}\left[\exp\left(-\frac{2}{(y - \tau)^\beta} - \left[\frac{2 + \alpha_l((U_{l, t+l} - U_{lt} - 1) + \max[U_{lt}, (1 - U_{l, t+l}]))}{y}\right]\right)\right] \\
 &= \mathbb{E}\left[\exp\left(-\frac{2}{(y - \tau)^\beta} - \left[\frac{2 - \alpha_l(\min[U_{lt}, (1 - U_{l, t+l}]))}{y}\right]\right)\right].
 \end{aligned}$$



**Figure A4.** Simulation Result 2.2: Graph of MCMC results for all parameters of model (5), showing time plot, estimated posterior distribution and autocorrelation.

(A.10) equals

$$\begin{aligned}
&= \frac{2 \left( 1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right) \right) - \left[ 1 - \mathbb{E} \left[ \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} + \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right] \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)} \\
&= 2 - \frac{\mathbb{E} \left[ 1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} + \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)} \\
&= 2 - \frac{1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} \right) \mathbb{E} \left[ \exp \left( \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)}.
\end{aligned} \tag{A.11}$$

$$\begin{aligned}
&\mathbb{E} \left[ \exp \left( \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right] = \mathbb{E} \left[ \exp \left( \frac{\alpha_l (\min [U_1, U_2])}{y} \right) \right] \\
&= \int_0^1 \int_0^{u_1} \exp \left( \frac{\alpha_l u_2}{y} \right) du_2 du_1 + \int_0^1 \int_0^{u_2} \exp \left( \frac{\alpha_l u_1}{y} \right) du_1 du_2 \\
&= 2 \frac{y^2}{\alpha_l^2} \left[ \exp \left( \frac{\alpha_l}{y} \right) - \frac{\alpha_l}{y} - 1 \right].
\end{aligned}$$

Therefore,

$$\rho_l(x) = P(X_{t+l} > x | X_t > x) = 2 - \frac{1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} \right) 2 \frac{y^2}{\alpha_l^2} \left[ \exp \left( \frac{\alpha_l}{y} \right) - \frac{\alpha_l}{y} - 1 \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)}.$$

**A.2.4. Proofs of Propositions 2.1 and 2.2.** Using Equation (A.11), for  $1 \leq l \leq L$  and  $y = (x/c)^\psi$ .

$$\begin{aligned}
\lambda_l &= \lim_{x \rightarrow \infty} P(X_{t+l} > x | X_t > x) = \lim_{y \rightarrow \infty} P(Y_{t+l} > y | Y_t > y) \\
&= \lim_{y \rightarrow \infty} 2 - \frac{\mathbb{E} \left[ 1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} + \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)}.
\end{aligned}$$

From Taylor's expansion and DCT, as  $y \rightarrow \infty$ .Case 1:  $\beta > 1$ ,

$$\begin{aligned}
&2 - \frac{\mathbb{E} \left[ 1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} + \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)} \\
&= 2 - \frac{\mathbb{E} \left[ \frac{2}{y} - \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right] + O \left( \frac{1}{y} \right)}{\frac{1}{y} + O \left( \frac{1}{y} \right)} \approx \mathbb{E} \left[ \alpha_l (\min [U_{lt}, (1-U_{l,t+l}]) \right].
\end{aligned}$$

Case 2:  $\beta = 1$ ,

$$\begin{aligned}
&2 - \frac{\mathbb{E} \left[ 1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} + \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right) \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)} \\
&= 2 - \frac{\mathbb{E} \left[ \frac{4}{y} - \frac{\alpha_l (\min [U_{lt}, (1-U_{l,t+l}])}{y} \right] + O \left( \frac{1}{y} \right)}{\frac{2}{y} + O \left( \frac{1}{y} \right)} \approx \frac{\mathbb{E} \left[ \alpha_l (\min [U_{lt}, (1-U_{l,t+l}]) \right]}{2}.
\end{aligned}$$

Case 3:  $\beta < 1$ ,

$$\begin{aligned}
 & 2 - \frac{\mathbb{E} \left[ 1 - \exp \left( -\frac{2}{(y-\tau)^\beta} - \frac{2}{y} + \frac{\alpha_t (\min[U_t, (1-U_{t,t+1})])}{y} \right) \right]}{1 - \exp \left( -\frac{1}{(y-\tau)^\beta} - \frac{1}{y} \right)} \\
 &= 2 - \frac{\mathbb{E} \left[ \frac{2}{y^\beta} \right] + O \left( \frac{1}{y^\beta} \right)}{\frac{1}{y^\beta} + O \left( \frac{1}{y^\beta} \right)} \approx 0.
 \end{aligned}$$