On the Estimation and Application of Max-Stable Processes

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

Richard L. Smith
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260
USA

August 22, 2008

Abstract

The theory of max-stable processes generalizes traditional univariate and multivariate extreme value theory by allowing for processes indexed by a time or space variable. We consider a particular class of max-stable processes, known as M4 processes, that are particularly well adapted to modeling the extreme behavior of multiple time series. We develop procedures for determining the order of an M4 process and for estimating the parameters. To illustrate the methods, some examples are given for modeling jumps in returns in multivariate financial time series. We introduce a new measure to quantify and predict the extreme co-movements in price returns.

Keywords: multivariate extremes, multivariate maxima of moving maxima, extreme value distribution, empirical distribution, estimation, extreme dependence, extreme co-movement.

1 Introduction

Extreme value theory is by now well established as a statistical technique for modeling data in which there is a particular interest in probabilities of very large or very small values. References such as Leadbetter, Lindgren and Rootzén (1983), Embrechts, Klüppelberg and Mikosch (1997), Coles (2001) and Smith (2003) have surveyed the theory, and there are many applications to environmental extremes, insurance and finance, amongst many other fields.

Multivariate extreme value theory is less widely used in practice, but there is still a substantial statistical theory and literature, see e.g. de Haan and Resnick (1977), Deheuvels (1978), Pickands (1981), de Haan (1985), Coles and Tawn (1991, 1994) and recent books by Beirlant *et al.* (2004), de Haan and Ferreira (2006) and Resnick (2007). These references are all based on the traditional definition of multivariate extremes, under which maxima or minima are defined componentwise across a sequence of random vectors. Alternative formulations such as those of Ledford and Tawn (1996, 1997) and Heffernan and Tawn (2004) will not be considered in the present paper. Applications of multivariate extreme value theory include all problems where there are several variables or processes being studied, and where an extreme value in any one of these is of interest. As one example, we could mention financial data sets where a large price change in any one of several assets could be of critical importance in managing a portfolio.

Max-stable processes are an infinite-dimensional generalization of multivariate extreme value theory that is particularly applicable in a time series or spatial process context. The applications of particular interest in the present paper are multiple financial time series where there is dependence in time as well as across the series. The mathematical theory was first laid out by de Haan (1984) and has been developed by a number of authors, e.g. Coles (1993), Schlather (2002), de Haan and Lin (2003), but applications have been relatively limited, despite a few attempts such as Coles and Tawn (1996).

The purpose of this paper is to develop some statistical theory for a particular class of max-stable process, known as *multivariate maxima of moving maxima*, or M4 processes for short (Smith and Weissman 1996). Although these processes have attractive probabilistic properties, there are difficulties in applying standard statistical estimation methods such as maximum likelihood. The general definition of a max-stable process involves a measure known as the spectral measure (de Haan, 1984) and M4 processes are a subclass for which the spectral measure is discrete. However, one consequence of this discreteness is that it implies certain degeneracies in the process itself, that we call *signature patterns*. It is unlikely that such signature patterns would be observed exactly in real data, though it is quite plausible that we would observe approximate signature patterns perturbed by random noise. Therefore, we would like an estimation method that is robust to such perturbations.

This difficulty has been noted previously for a class of processes closely related to M4 processes, the so-called *max-autoregressive moving average* or MARMA processes of Davis and Resnick (1989, 1993). Hall, Peng and Yao (2002) got around this difficulty by defining a class of estimators based on empirical processes. The method proposed here for M4 processes is similarly motivated.

The paper is organized as follows. In Section 2, we introduce the M4 process and list some key properties. The estimators and their asymptotic properties are studied in Section 3. In contrast to the bootstrapped processes which Hall *et al.* (2002) used to construct confidence intervals and prediction intervals for moving maxima models, we directly construct parameter estimators and prove their asymptotic properties for the M4 processes. In Section 4 we provide simulation examples to show the efficiency of proposed estimating procedure. In Section 5 we explore modeling financial time series data as M4 processes. Returns of spot exchange rate of Japanese Yen against US dollar (JPY/USD), Canadian dollar against US dollar (CAD/USD), and British pound against US dollar (GBP/USD) are studied. Section 6 contains discussion and conclusions. Section 7 contains details of some of the algebraic derivations which, because of their complexity, are not fully spelled out in the main text. Section 8 contains detailed technical proofs.

2 The model and its identifiability

Our starting point is a multivariate strongly stationary time series $\{X_{id},\ i = 0, \pm1, \pm2, ..., d = 1, ..., D\}$, where i is time and d indexes a component of the process.

A standard method of univariate extreme value theory is to model the exceedances above a high threshold by the generalized Pareto distribution (Davison and Smith, 1990). Assuming this and applying a probability integral transformation, it is possible to transform each marginal distribution of the process, above a high threshold, so that the marginal distribution is unit Fréchet. For the moment we ignore the high threshold part of the modeling, and assume that the univariate Fréchet assumption applies to the whole distribution. Thus, we transform each X_{id} into a random variable Y_{id} for which $\Pr\{Y_{id} \leq y\} = \exp(-1/y),\ 0 < y < \infty$.

The process $\{Y_{id}\}$ is said to be max-stable if for any finite collection of time points $i = i_1, i_1 + 1, ..., i_2$ and any positive set of values $\{y_{id},\ i = i_1, ..., i_2,\ d = 1, ..., D\}$, we have

$$\Pr\{Y_{id} \leq y_{id},\ i = i_1, ..., i_2,\ d = 1, ..., D\} = \left[\Pr\{Y_{id} \leq n y_{id},\ i = i_1, ..., i_2,\ d = 1, ..., D\}\right]^n .$$

This property directly generalizes the max-stability property of univariate and multivariate extreme value distributions (Leadbetter *et al.* (1983), Resnick (1987)) and provides a convenient mathematical framework to talk about extremes in infinite-dimensional processes.

Smith and Weissman (1996) proved the following characterization of max-stable processes: under some mixing assumptions that we shall not detail here, any max-stable process with unit Fréchet margins may be approximated by a *multivariate maxima of moving maxima* process, or M4 for short, with the representation

$$Y_{id} = \max_{l=1,2,\ldots} \max_{-\infty<k<\infty} a_{l,k,d} Z_{l,i-k},\ -\infty < i < \infty,\ d = 1, ..., D,$$

where $\{Z_{l,i},\ l = 1, 2, ..., -\infty < i < \infty\}$ are independent unit Fréchet random variables and $a_{l,k,d}$ are non-negative coefficients satisfying $\sum_{l=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{l,k,d} = 1$ for each d.

In practice, even this representation is too cumbersome for practical application, involving infinitely many parameters $a_{l,k,d}$, so we simplify it by assuming that only a finite number of these coefficients are non-zero. Thus we have the representation

$$Y_{id} = \max_{1\leq l\leq L_d} \max_{-K_{1ld}\leq k\leq K_{2ld}} a_{l,k,d} Z_{l,i-k},\ -\infty < i < \infty, \quad d = 1, \ldots, D, \tag{2.1}$$

where $L_d,\ K_{1ld},\ K_{2ld}$ are finite and the coefficients satisfy $\sum_{l=1}^{L_d} \sum_{k=-K_{1ld}}^{K_{1ld}} a_{l,k,d} = 1$ for each d.

Probabilistic properties of the model (2.1) have been studied in Zhang and Smith (2004), Martins and Ferreira (2005), among others. We have a general joint probability formula:

$$\Pr\{Y_{id} \leq y_{id}, 1 \leq i \leq r, 1 \leq d \leq D\} = \exp\Big[-\sum_{l=1}^{\max_d L_d} \sum_{m=1-\max_d K_{2ld}}^{r+\max_d K_{1ld}} \max_{1-m\leq k\leq r-m} \max_{1\leq d\leq D} \frac{a_{l,k,d}}{y_{m+k,d}}\Big], \tag{2.2}$$

where $a_{l,k,d} = 0$ when the triple subindex is outside the range defined in (2.1). This assumption is held in the rest of the paper. Besides this general formula, it follows immediately that $\Pr(Y_{id} \leq y) = e^{-1/y}$, which establishes that Y_{id} is itself a unit Fréchet random variable, and the following two special cases which are used extensively in the subsequent discussion:

$$\Pr(Y_{id} \leq y_{id}, Y_{i+1,d} \leq y_{i+1,d}) = \exp\Big[-\sum_{l=1}^{L_d} \sum_{m=1-K_{2ld}}^{2+K_{1ld}} \max\Big\{\frac{a_{l,1-m,d}}{y_{id}}, \frac{a_{l,2-m,d}}{y_{i+1,d}}\Big\}\Big], \tag{2.3}$$

$$\Pr(Y_{id} \leq y_{1d}, Y_{id'} \leq y_{1d'}) = \exp\Big[-\sum_{l=1}^{\max(L_d,L_{d'})} \sum_{m=1-\max(K_{2ld},K_{2ld'})}^{1+\max(K_{1ld},K_{1ld'})} \max\Big\{\frac{a_{l,1-m,d}}{y_{1d}}, \frac{a_{l,1-m,d'}}{y_{1d'}}\Big\}\Big]. \tag{2.4}$$

We also note the following, which is used several times later:

Remark 1 *For each l and d, the value of $\sum_{k=-K_{1ld}}^{K_{2ld}} a_{l,k,d}$ is the asymptotic proportion (as $u \to \infty$) of the total number of clusters of exceedances of a high threshold u by the d'th component process that are drawn from the lth signature pattern which will be defined next.*

Under model (2.1), an extremal event is typically associated with a single very large value of Z, say Z_{lk}. When this happens, we will have $Y_{id} = a_{l,i-k,d}Z_{lk}$ for several values of i near k. Thus the sequence $\{Y_{id}\}$ is locally proportional to a deterministic sequence $a_{l,i-k,d}$ for some $l \in \{1, \ldots, L_d\}$, which we call a signature pattern. Here L_d corresponds to the maximum number of distinct signature patterns. The constants K_{1ld} and K_{2ld} characterize the range of dependence in each sequence and $(\max_l K_{1ld} + \max_l K_{2ld} + 1)$ is the order of the moving maxima processes. We illustrate these phenomena in Figure 1. Plot (a) shows a simulated sample of length 365 from an M4 process. Plots (b) and (c) are blown up plots of parts of the series near local maxima at $i = 42$ and 103 respectively. It can be seen that, even though the y-scales of the two plots are quite different, the shapes are identical. Plots (d) and (e) illustrate the same phenomenon but where a different value of l is responsible for the shape. These characteristic shapes are known as signature patterns.

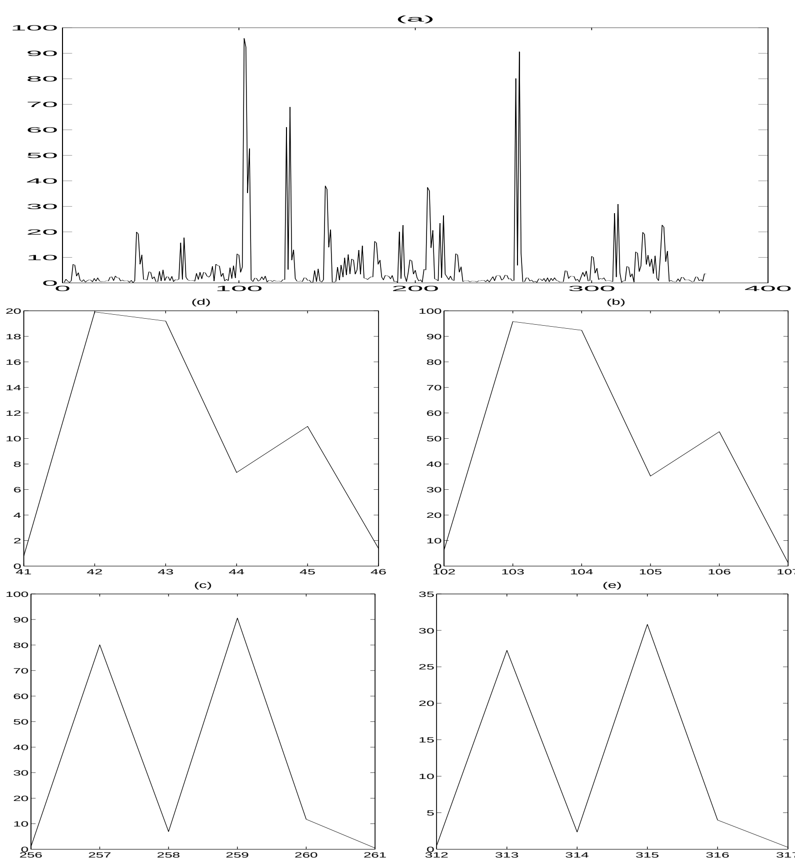

Figure 1: *A demonstration of an M4 process. (a) is a simulated 365 days data for a component process. (b) - (e) are partial pictures drawn from the whole simulated data showing two different moving patterns, called signature patterns, in certain time periods when extremal events occur.*

In real data, it is unlikely that we would observe exact replicates of a signature pattern as in Figure

1. However, the extremal properties of an M4 process are not changed by short-tailed additive noises. For the robustness of the regular variation property with respect to short-tailed additive noises, we refer to the appendix in Embrechts *et al* (1997). We present a proposition which shows that the asymptotic dependence properties do not change with additive noises in an M4 process.

For any pair of random variables Y_1 and Y_2 sharing the same marginal distribution with common upper endpoint $y_F \leq \infty$, we may define the *bivariate tail dependence index* (Sibuya, 1960) to be

$$\lambda = \lim_{u \to y_F} \Pr(Y_1 > u | Y_2 > u). \tag{2.5}$$

Under Model (2.1), the tail dependence index $\lambda_{dd'}$ between Y_{id} and $Y_{id'}$ is:

$$\lambda_{dd'} = 2 - \sum_{l=1}^{\max(L_d, L_{d'})} \sum_{m=1-\max(K_{2ld}, K_{2ld'})}^{1+\max(K_{1ld}, K_{1ld'})} \max\left\{a_{l,1-m,d}, a_{l,1-m,d'}\right\}.$$

Moreover, for fixed d, $Y_{i_1 d}$ and $Y_{i_2 d}$ have positive tail dependence whenever $|i_1 - i_2| \leq k = \max_{1 \leq l \leq L_d}(K_{1ld} + K_{2ld})$, and are therefore said to be lag-k tail dependent. The tail dependence index between $Y_{i_1 d}$ and $Y_{i_2 d}$, when $|i_2 - i_1| = k$, is

$$\lambda_{d(k)} = 2 - \sum_{l=1}^{L_d} \sum_{m=1-K_{2ld}}^{1+k+K_{1ld}} \max\left\{a_{l,1-m,d}, a_{l,1+k-m,d}\right\}.$$

Proposition 2.1 *Suppose X and Y are asymptotically dependent unit Fréchet random variables, i.e. $\lambda = \lim_{u \to \infty} \Pr(Y > u | X > u) > 0$. Suppose $P(|N_i| > u) = o(u^{-1})$, $i = 1, 2$, as $u \to \infty$. Then $X^* = X + N_1$ and $Y^* = Y + N_2$ are asymptotically dependent with the same tail dependence index λ.*

The proof of Proposition 2.1 is elementary and therefore omitted.

An M4 process with noise is much more realistic for practical time series than an unperturbed M4 process, and is the focus of our subsequent methodology. However, for this reason, we do not apply the method of maximum likelihood, but instead propose estimators based on using (2.3) - (2.4) to approximate empirical bivariate distribution functions.

The idea of estimating the $\{a_{l,k,d}\}$ parameters from bivariate distributions naturally raises the question of whether bivariate distributions identify all the parameters of the process. In the following discussion we propose sufficient, though not necessary, conditions for that.

The probability evaluated at the points $(y_{id}, y_{i+1,d})$ in (2.3) depends on the comparison of $a_{l,1-m,d}/y_{id}$ and $a_{l,2-m,d}/y_{i+1,d}$, and similarly in (2.4). By fixing one of y_{id} and $y_{i+1,d}$ — say y_{id} — then $a_{l,1-m,d}/(a_{l,2-m,d} y_{id})$ is the change point of $\max(a_{l,1-m,d}/y_{id}, a_{l,2-m,d}/y_{i+1,d})$ when $y_{i+1,d}$ varies. By max-stability, we immediately have

$$\Pr\{Y_{id} \leq u, Y_{i+1,d} \leq u + x\} = \Pr\{Y_{id} \leq 1, Y_{i+1,d} \leq (u+x)/u\}^{(1/u)}.$$

So without loss of generality, we can fix $y_{id} = 1$. In a real data application we might choose a threshold value u and fix $y_{id} = u$.

From (2.3) and (2.4), we have

$$\Pr\{Y_{id} \le 1, Y_{i+1,d} \le x\} = \exp\Big[-\sum_{l=1}^{L_d} \sum_{m=1-K_{2ld}}^{2+K_{1ld}} \max(a_{l,1-m,d}, \frac{a_{l,2-m,d}}{x}) \Big],$$

and

$$\Pr\{Y_{id} \le 1, Y_{id'} \le x\} = \exp\Big[- \sum_{l=1}^{\max(L_d,L_{d'})} \sum_{m=1-\max(K_{2ld},K_{2ld'})}^{1+\max(K_{1ld},K_{1ld'})} \max\big(a_{l,1-m,d}, \frac{a_{l,1-m,d'}}{x}\big) \Big].$$

Let

$$b_d(x) = -\log\big[\Pr(Y_{id} \le 1,\ Y_{i+1,d} \le x)\big],\ b_{dd'}(x) = -\log\big[\Pr(Y_{id} \le 1, Y_{id'} \le x)\big].$$

We have

$$\begin{aligned} b_d(x) \quad = \quad & \sum_{l=1}^{L_d} \Big[a_{l,K_{2ld},d} + \max(a_{l,K_{2ld}-1,d}, \frac{a_{l,K_{2ld},d}}{x}) + \max(a_{l,K_{2ld}-2,d}, \frac{a_{l,K_{2ld}-1,d}}{x}) \\ & + \max(a_{l,K_{2ld}-3,d}, \frac{a_{l,K_{2ld}-2,d}}{x}) + \cdots + \max(a_{l,-K_{1ld},d}, \frac{a_{l,-K_{1ld}+1,d}}{x}) + \frac{a_{l,-K_{1ld},d}}{x} \Big], \\ & \quad d = 1, \ldots, D, \end{aligned} \tag{2.6}$$

and

$$b_{dd'}(x) = \sum_{l=1}^{\max(L_d,L_{d'})} \sum_{m=1-\max(K_{2ld},K_{2ld'})}^{1+\max(K_{1ld},K_{1ld'})} \max(a_{l,1-m,d},\ \frac{a_{l,1-m,d'}}{x}). \tag{2.7}$$

It is clear that for each d, we can define new piecewise linear functions: $q_d(x) \triangleq xb_d(x)$ and $q_{dd'}(x) \triangleq xb_{dd'}(x)$, where the notation $A \triangleq B$ means that A is denoted as B, and the points where these piecewise linear functions change slopes are at $a_{l,j,d}/a_{l,j',d}$ or $a_{l,k,d}/a_{l,k,d'}$. A typical $q_d(x)$ picture is shown in Figure 2 which is drawn for a particular model that is defined in Example 2.1 which is presented at the end of this section. This suggests that if we can identify the functions $q_d(x)$ or $q_{dd'}(x)$, we may be able to identify all the parameters $a_{l,k,d}$.

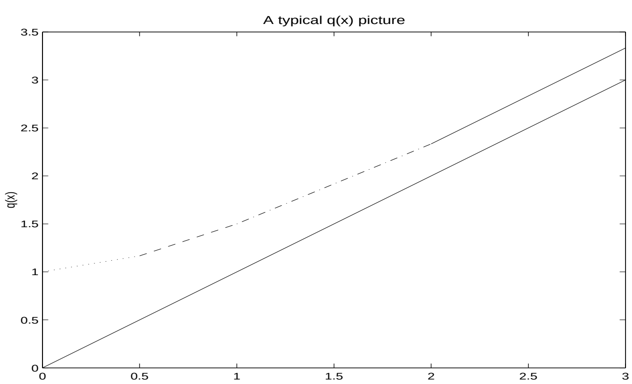

Figure 2: The function $q_d(x)$ defined for Example 2.1. This is a piecewise linear function with changes of slope at $x = 0.5$, 1 and 2.

However, in practice, we do not want to evaluate these functions on a continuum of x values, and it would be much more convenient if we could get away with evaluating them for a finite set of xs.

It turns out we can do that and still retain identifiability, provided that the finite set of xs is chosen appropriately. The following important Proposition 2.2 will show the identifiability. It plays a core mathematical role in parameter estimation. Its proof is nonstandard and is deferred to Section 8.

Proposition 2.2 *Suppose the following three conditions hold:*

(i) all nonzero and existing ratios $\frac{a_{l,j,d}}{a_{l,j',d}}$ *for all* l *and* $j \neq j'$ *are distinct for each* $d = 1, \ldots, D$*;*

(ii) for all $l \neq l'$ *and,* $j,\ j'$, $a_{l,j,d} \neq a_{l',j',d}$ *when both are greater than zero for each* $d = 1, \ldots, D$*;*

(iii) and nonzero and existing ratios $\frac{a_{l,k,1}}{a_{l',k,d'}}$ *for all* $l,\ l'$ *and* k *are distinct for each* $d' = 2, \ldots, D$*;*

then $b_d(x),\ d = 1, \ldots, D,\ b_{1d'}(x),\ d' = 2, \ldots, D$ *uniquely determine all values of* $a_{l,k,d},\ d = 1, \ldots, D,\ l = 1, \ldots, L_d,\ -K_{1ld} \leq k \leq K_{2ld}$.

Furthermore, there exist points $x_1, x_2, \ldots, x_m$, $m \leq (2D-1)\sum_d \sum_l^{L_d}(K_{1ld}+K_{2ld}+2)+2D$*, such that*

$$b_d(x_i) \quad and \quad b_{1d'}(x_i), i = 1, \ldots, m,\ d = 1, \ldots, D,\ d' = 2, \ldots, D$$

uniquely determine all values of $a_{l,k,d}$.

Remark 2 *When* $D = 1$ *and* $L_d = 1$*, the reason why Proposition 2.2 is true is that in this case, any permutation of the* a_{jd}*'s must create a new set of values of ratios or jump points which will result in a different function of* $q_d(x)$*. Condition (ii) assures that any permutation between different signature patterns will give a different* $b_d(.)$ *function. Condition (iii) combines* D *univariate processes into a unique joint M4 process.*

This justifies statements like "for almost all (w.r.t Lebesgue measure) choices of coefficients $a_{-K_{1ld},d}, \ldots, a_{K_{2ld},d}$*, the model is identifiable from* $q_d(x)$*."*

Remark 3 *Only* $b_{1d'}(x)$ *is used in the proof of model identifiability. In some situations, other* $b_{dd'}(x)$ *functions may also be needed in order to prove identifiability or to get estimates of all parameters. Also, notice that when we have some* a_{lkd}*s being zero for some* $-K_{1ld} < k < K_{2ld}$*, some of the ratios may be infinite, and not all coefficients* a_{lkd} *are identifiable by function* $b_d(x)$*. This issue can be resolved by simply replacing* $b_d(x)$ *by* $b_{d(r)}(x) = -\log\left[\Pr(Y_{id} \leq 1,\ Y_{i+r,d} \leq x)\right]$ *for an appropriately chosen* r*, where* $b_{d(1)}(x)$ *is just* $b_d(x)$ *in (2.6). However, these changes do not make any difference to the format of the results.*

Although we believe that the conditions of Proposition 2.2 are general enough to cover most practical cases, there are cases where the conditions do not hold, for which bivariate distributions are insufficient to determine all the coefficients. Here is an example.

Example 2.1 *Let* $(a_0, \ldots, a_4) = \frac{1}{6}(1, 1, 2, 1, 1)$ *and* $(b_0, \ldots, b_4) = \frac{1}{6}(1, 2, 1, 1, 1)$*. We consider the two processes generated by the sequences* $a_0, \ldots, a_4$ *and* $b_0, \ldots, b_4$*, i.e.* $Y_i = \max_{k=0,1,2,3,4} a_k Z_{i-k},\ -\infty < i < \infty$ *and* $Y_i' = \max_{k=0,1,2,3,4} b_k Z_{i-k},\ -\infty < i < \infty$*. For either process, the function* $q(x)$ *(Figure 2) is piecewise linear with changepoints at* $r_1 = \frac{1}{2}, r_2 = 1, r_3 = 2$ *and slopes that are respectively* $\frac{1}{6}, \frac{1}{2}, \frac{5}{6}, 1$ *on* $(0, \frac{1}{2}), (\frac{1}{2}, 1), (1, 2), (2, \infty)$*. Therefore, the two processes are indistinguishable from the bivariate distributions. However, we also have*

$$\begin{aligned} -\log(\Pr(Y_1 \leq y_1, Y_2 \leq y_2, Y_3 \leq y_3)) \quad = \quad & \tfrac{a_4}{y_1} + \max(\tfrac{a_3}{y_1}, \tfrac{a_4}{y_2}) + \max(\tfrac{a_2}{y_1}, \tfrac{a_3}{y_2}, \tfrac{a_4}{y_3}) + \max(\tfrac{a_1}{y_1}, \tfrac{a_2}{y_2}, \tfrac{a_3}{y_3}) \\ & + \max(\tfrac{a_0}{y_1}, \tfrac{a_1}{y_2}, \tfrac{a_2}{y_3}) + \max(\tfrac{a_0}{y_2}, \tfrac{a_1}{y_3}) + \tfrac{a_0}{y_3} \end{aligned}$$

so if $y_1 = 1, y_2 = y_3 = c > 2$, *we have* $-\log(\Pr(Y_1 \le y_1, Y_2 \le y_2, Y_3 \le y_3)) = 1 + \frac{1}{3c}$ *but* $-\log(\Pr(Y_1' \le y_1, Y_2' \le y_2, Y_3' \le y_3)) = 1 + \frac{1}{2c}$, *so the trivariate distributions are different.*

3 The estimators and asymptotics

We now propose estimators for parameters in general M4 processes. The basic idea behind the estimation technique is to estimate empirical bivariate distribution functions and then solve for the $a_{l,k,d}$ coefficients. The empirical counterparts of $b_d(x)$ and $b_{1d'}(x)$ are defined as:

$$U_d(x) = \frac{1}{n}\sum_{i=1}^{n-1} I_{\{Y_{id}\le 1,\ Y_{i+1,d}\le x\}},\ \widehat{b}_d(x) = -\log\big[U_d(x)\big],\ d=1,\ldots,D, \tag{3.1}$$

$$U_{1d'}(x) = \frac{1}{n}\sum_{i=1}^{n-1} I_{\{Y_{i1}\le 1,\ Y_{id'}\le x\}},\ \widehat{b}_{1d'}(x) = -\log\big[U_{1d'}(x)\big],\ d'=2,\ldots,D, \tag{3.2}$$

where $I(.)$ is an indicator function. Let

$$x_{1d},\ x_{2d},\ \ldots,\ x_{md},\ d=1,\ \ldots,\ D,$$

where $m \le (2D-1)(\max_l K_{1ld} + \max_l K_{2ld} + 1) + 2D$ as described in Proposition 2.2, and

$$x'_{1d'},\ x'_{2d'},\ \ldots,\ x'_{m'd'},\ d'=2,\ \ldots,\ D,$$

where $m' \le (2D-1)\sum_d \sum_l^{L_d}(K_{1ld}+K_{2ld}+2)+2D$, be suitable choices of the points used to evaluate the functions. Then (3.1) and (3.2) can be written as the following vector forms:

$$\mathbf{x} = \Big(x_{11},\ x_{21},\ \ldots,\ x_{m1},\ x_{12},\ \ldots,\ x_{mD},\ x'_{12},\ x'_{22},\ \ldots,\ x'_{m'2},\ x'_{13},\ \ldots,\ x'_{m'D}\Big)^T,$$

$$\mathbf{U} = \Big(U_1(x_{11}),\ \ldots,\ U_1(x_{m1}),\ U_2(x_{12}),\ \ldots,\ U_D(x_{mD}),\ U_{12}(x'_{12}),\ \ldots,\ U_{12}(x'_{m'2}),\ \ldots,\ U_{1D}(x'_{m'D})\Big)^T,$$

$$\widehat{\mathbf{b}} = \Big(\widehat{b}_1(x_{11}),\ \ldots,\ \widehat{b}_1(x_{m1}),\ \widehat{b}_2(x_{12}),\ \ldots,\ \widehat{b}_D(x_{mD}),\ \widehat{b}_{12}(x'_{12}),\ \ldots,\ \widehat{b}_{12}(x'_{m'2}),\ \ldots,\ \widehat{b}_{1D}(x'_{m'D})\Big)^T.$$

In the literature, the central limit theorem for m-dependent sequences is well established, for example, Berk (1975), Billingsley (1995), among others. Because of the complexity of some of the definitions, the calculation of asymptotic covariance matrix is not simple in this study, and we do not give them all here but refer to Section 7 for full details. One of our main results is:

Lemma 3.1 *For the choices of* $x_{jd},\ x_{j'd'}$ *and with* $\mu,\ \mathbf{b},\ \Sigma,\ W_k,\ \Theta$ *defined in Section 7, we have*

$$\sqrt{n}(\mathbf{U}-\boldsymbol{\mu}) \xrightarrow{\mathcal{L}} N\Big(0, \Sigma + \sum_{k=1}^{\max_l K_{1ld}+\max_l K_{2ld}+1}\{W_k + W_k^T\}\Big),$$

and

$$\sqrt{n}(\widehat{\mathbf{b}}-\mathbf{b}) \xrightarrow{\mathcal{L}} N\Big(0, \Theta\big[\Sigma + \sum_{k=1}^{\max_l K_{1ld}+\max_l K_{2ld}+1}\{W_k + W_k^T\}\big]\Theta^T\Big),$$

which establish the asymptotics for the empirical functions $\widehat{b}_d(x),\ \widehat{b}_{1d'},\ d=1,\ldots,D,\ d'=2,\ldots,D.$

Proof of Lemma 3.1 is deferred to Section 8.

The results in Lemma 3.1 are for any arbitrary choices of $x_{jd},\ x_{j'd'}$. In order to construct estimators in M4 models, these choices must satisfy the identifiability conditions discussed in the previous section. For the moment we assume these conditions are satisfied — how they are determined in practice is discussed in Section 4.

The next step in the estimation procedure is that we use the estimated $\widehat{\mathbf{b}}$ to solve for the $a_{l,k,d}$ values. Consider the system of non-linear equations formed by (2.6) and (2.7), where x is now substituted by $x_{jd},\ j=1,\ldots,m,\ d=1,\ldots,D$ in (2.6), and by $x'_{j'd'},\ j'=1,\ldots,m',\ d'=2,\ldots,D$ in (2.7), respectively. The left hand sides of these equations collectively define the vector $\mathbf{b}$, in the same notation as Lemma 3.1. Let $\mathbf{a}$ denote the vector whose elements are all parameters $a_{l,k,d}$. Since (2.6) and (2.7) uniquely determine the values of all parameters $a_{l,k,d}$, each of the maxima in (2.6) and (2.7) is determined uniquely (no ties). Therefore, the relation between $\mathbf{b}$ and $\mathbf{a}$ evaluated based on Equations (2.6) and (2.7) has the matrix representation

$$\mathbf{b} = C\mathbf{a}, \tag{3.3}$$

where each element in matrix C belongs to $\{1,\ 1/x_{jd},\ 1+1/x_{jd},\ 1/x'_{j'd'},\ j=1,\ldots,m,\ d=1,\ldots,D,\ j'=1,\ldots,m',\ d'=2,\ldots,D\}$, and C^TC is invertible. The following Lemma 3.2 is important in proving our main Theorem 3.3.

Lemma 3.2 *Suppose S is a set with finite number of distinct values, $\mathbf{b}^*$ and $\mathbf{a}^*$ are vectors in $\mathbb{R}^m$. Suppose C^* is an $l\times m$ matrix whose elements belong to S, $C^{*T}C^*$ is invertible, and C^* is the unique matrix such that $\mathbf{b}^*=C^*\mathbf{a}^*$. Suppose $\{\mathbf{b}_n^*, n=1,2,\ldots\}$ is a sequence of random vectors, $\{\mathbf{a}_n^*, n=1,2,\ldots\}$ is a sequence of random vectors, and $\{C_n^*, n=1,2,\ldots\}$ is a sequence of random matrices satisfying $C_n^{*T}C_n^*$ being invertible. Suppose $\mathbf{b}_n^*=C_n^*\mathbf{a}_n^*,\ n=1,2,\ldots$. $\mathbf{b}_n^*\xrightarrow{a.s.}\mathbf{b}^*,\ C_n^*\mathbf{a}_n^*\xrightarrow{a.s.}C^*\mathbf{a}^*$, as $n\to\infty$, then $C_n^*\xrightarrow{a.s.}C^*$, $\mathbf{a}_n^*\xrightarrow{a.s.}\mathbf{a}^*$, as $n\to\infty$, and $||C_n^*-C^*||=o_p(1)$.*

Proof of Lemma 3.2 is deferred to Section 8.

From the estimate $\widehat{\mathbf{b}}$, and (2.6) and (2.7), we have

$$\widehat{\mathbf{b}} = \widehat{C}\widehat{\mathbf{a}}, \tag{3.4}$$

where each element in matrix $\widehat{C}$ belongs to $\{1,\ 1/x_{jd},\ 1+1/x_{jd},\ 1/x'_{j'd'},\ j=1,\ldots,m,\ d=1,\ldots,D,\ j'=1,\ldots,m',\ d'=2,\ldots,D\}$, which is a finite set. Since the estimators obey the strong law of large numbers, $\widehat{\mathbf{b}}$ converges to $\boldsymbol{b}$ as $n\to\infty$, and hence (by Lemma 3.2) for sufficiently large n, we therefore have the representation:

$$\widehat{\mathbf{b}} = C\widehat{\mathbf{a}},$$

which is equivalent to

$$(C^TC)^{-1}C^T\widehat{\mathbf{b}} = \widehat{\mathbf{a}}. \tag{3.5}$$

Summarizing all arguments above, we have obtained the following theorem which is the asymptotic distribution of the estimators.

Theorem 3.3 *For the multivariate processes $\{Y_{id}\}$, suppose all three conditions in Proposition 2.2 are satisfied, then there exist*

$$\{x_{1d},\ x_{2d},\ \ldots,\ x_{md},\ d=1,\ \ldots,\ D\},$$

and

$$\{x'_{1d'},\ x'_{2d'},\ \ldots,\ x'_{m'd'},\ d'=2,\ \ldots,\ D\},$$

such that the estimator $\widehat{\mathbf{a}}$*, which is the solution of (3.5), satisfies*

$$\sqrt{n}(\widehat{\mathbf{a}}-\mathbf{a}) \xrightarrow{\mathcal{L}} N\Big(0, B\Theta\big[\Sigma + \sum_{k=1}^{\max_{l,d} K_{1ld}+\max_{l,d} K_{2ld}+1} \{W_k + W_k^T\}\big]\Theta^T B^T\Big)$$

where $B=(C^TC)^{-1}C^T$.

For some examples of M4 processes, Theorem 3.3 is directly applicable. In others, however, we find that the conditions of Theorem 3.3 are applicable to any of the 1-dimensional component processes, but not to the joint distributions of the D-dimensional process. A concrete example is in Section 5. For such cases, we propose an alternative methodology and use an example to illustrate the idea.

Example 3.1 *Suppose, for the dth component process on its own, the coefficients are* $\{a^*_{l^*,k,d}\}$*. Typically for the full D-dimensional process, for each l and d we have that* $\{a_{l,k,d},\ k=-K_{1ld},...,K_{2ld}\}$ *are proportional to* $\{a^*_{l^*,k,d},\ k=-K_{1ld},...,K_{2ld}\}$ *for some* l^* *which is a function of* (l,d)*, but the number of distinct signature patterns in* $\{a^*_{l,k,d}\}$ *is smaller than in* $\{a_{l,k,d}\}$*. In this case we write* $a_{l,k,d}=\beta_{l,d}a^*_{l^*,k,d},\ \beta_{ld}\geq 0,\ a^*_{l^*,k,d}\geq 0$*, and the joint model is*

$$\begin{aligned} Y_{id} &= \max_{1\leq l\leq L}\ \max_{-K_1\leq k\leq K_2} a_{l,k,d}Z_{l,i-k} \\ &= \max_{1\leq l\leq L}\beta_{ld} \max_{-K_1\leq k\leq K_2} a^*_{l,k,d}Z_{l,i-k},\ d=1,\ldots,D \end{aligned} \tag{3.6}$$

where $(a^*_{l,k,d},\ k=-K_1,\ldots,K_2)=(a^*_{l',k,d},\ k=-K_1,\ldots,K_2)$ *for some* $l\neq l'$*, and* $\sum_{l\in S_{l^*d}}\beta_{ld}=1$*,* S_{l^*d} *contains all ls and l′s such that* $(a^*_{l,k,d},\ k=-K_1,\ldots,K_2)=(a^*_{l',k,d},\ k=-K_1,\ldots,K_2)$ *which is related to a particular* l^**.*

This example tells that the exact number of signature patterns in each component process may be less than L. Some observed signature patterns are matched with other component processes. We may understand this like: a particular signature pattern is split into several "signature patterns" based on relative proportion parameters β_{ld}. One can see in the above example that not all nonzero existing ratios $\frac{a_{l,j,d}}{a_{l,j',d}}$ are distinct. However, estimation of a_{lkd} can be done via estimating $a^*_{l^*kd}$ and β_{ld} first.

For the present discussion we simplify the problem a little by assuming the $\beta_{l,d}$ are known though in practice they would also have to be estimated (Section 5 gives a specific example). The result in this case is:

Corollary 3.4 *Suppose for each d in (3.6),* $a^*_{l,k,d}s$ *are estimated using observations from the dth component process. The limiting covariance matrices are denoted as* $\Sigma^*_1,\ \Sigma^*_2,\ldots,\Sigma^*_D$ *respectively. Suppose the vector* $\mathbf{a}^*$ *consists of all the coefficients* $a^*_{l,k,d}$ *arranged in some order, and the corresponding estimators* $\widehat{a}^*_{l,k,d}$ *are arranged in* $\widehat{\mathbf{a}}^*$ *accordingly, then*

1. $\sqrt{n}\,(\widehat{\mathbf{a}}^*-\mathbf{a}^*) \xrightarrow{\mathcal{L}} N(0,\mathbf{\Sigma^*})$*, where the diagonal matrices in* $\mathbf{\Sigma^*}$ *are* $\Sigma^*_1,\ \Sigma^*_2,\ldots,\Sigma^*_D$*.*

2. *There is a matrix* $\mathbf{\Sigma}$ *(depending on* $\mathbf{\Sigma^*}$ *and the coefficients* $\{\beta_{l,d}\}$*) so that*

$$\sqrt{n}\,(\widehat{\mathbf{a}} - \mathbf{a}) \xrightarrow{\mathcal{L}} N(0, \mathbf{\Sigma}).$$

The proof of this corollary is simply due to the fact that $\widehat{\mathbf{a}}$ coefficients are linear combinations of $\widehat{\mathbf{a}}^*$ which has an asymptotic multivariate normal distribution. The arrangement does not depend on the sample size n. The form of $\mathbf{\Sigma}$ depends on particular identical signature patterns in $\{a_{l,k,d}\}$. The asymptotic standard deviations of $\{\widehat{a}_{l,k,d}\}$ can be easily obtained from $\Sigma_1^*,\ \Sigma_2^*, \ldots, \Sigma_D^*$, and $\beta_{l,d}$s.

Theorem 3.3 and Corollary 3.4 establish the asymptotic distribution of the parameter estimators. In the next section we propose a procedure to determine the actual x_{jd} used in the estimating equations.

4 Determining the x_{jd} values and a simulation example

We propose using a standard cluster analysis method, such as k-means nearest neighbor clustering, to group very large clustered observations (above certain thresholds) into L_d groups based on the consecutive ratios of $Y_{i+j-1,d}/Y_{i+j,d},\ j = 1, ..., K$. The proposed procedure to determine the x_{jd} and $x'_{j'd}$ values is then as follows:

1. For each d, use cluster analysis to group the consecutive ratios of $(Y_{i+j-1,d}/Y_{i+j,d},\ j = 1, \ldots, K)$ into L_d groups for all very large clustered observations indexed on i and appearing in K consecutive days. The tuning parameters L_d and K are assumed known in this section.

2. For all clustered groups, assign the same group number to the cases where the consecutive ratios of $(Y_{i+j-1,d}/Y_{i+j,d},\ j = 1, \ldots, K)$ are in the same cluster. The group numbers across component processes also need to be the same when we observe signature patterns over several component processes simultaneously.

3. Within each group, take the averages of the ratios as points where the function $q_d(x)$ changes slope. Between any two adjacent points, arbitrarily choose two points as x_{jd} values. For example, suppose $r_1,\ r_2$ are two adjacent ratios, then a natural choice would be $x_{jd} = r_1 + .25(r_2 - r_1)$, $x_{j+1,d} = r_1 + .75(r_2 - r_1)$.

4. The choices of $x'_{j'd}$ can be done from averaging the ratios of Y_{i1}/Y_{id} within the same group numbers obtained in Step 2 between two processes. Then $x'_{j'd}$ can take the middle values of two adjacent ratios or take two values between two adjacent ratios like the previous step.

5. After choosing x_{jd} and $x'_{j'd}$ values, use them to estimate $a_{l,k,d}$ based on $\widehat{b}_d(x)$ and $\widehat{b}_{1d}(x)$ functions.

Remark 4 *In Step 1, we may need to cluster those very large clustered observations into more than* L_d *groups since outliers may exist and cause the clustering method to fail to recognize the true patterns. In our example, we cluster those observations into* $L_d + 3$ *groups. The 3 groups which have very small proportions among all those very large clustered observations are not used in determining* x_{jd} *values. We note that adding 3 groups works with this particular example. In other examples, we may need to add different number of groups in clustering analysis.*

Remark 5 *In Step 3 and 4, theoretically, we should choose as many points of x_{jd} and $x'_{j'd}$ as possible, but this is not realistic due to the intensive computation and the complexity of inferences. The goal is to choose moderate number of points such that the estimated values of parameters are close to the true parameter values. In our simulation study, we found two times the number of clustered group patterns works well based on the comparisons between estimated parameter values and their corresponding true values.*

As an illustration of these techniques, we show how they work on a simple simulated process that follows exactly the model of M4 plus noise. Suppose $D = 2$ and

$$Y_{id} = \max_{1\leq l\leq 3} \max_{-1\leq k\leq 1} a_{l,k,d} Z_{l,i-k} + N_{id}, \ -\infty < i < \infty, \quad d = 1, 2, \tag{4.1}$$

where each M4 process has three signature patterns and moving range order of 3, the noises $N_{id} \sim N(0, 1)$. The coefficients are listed in Table 1.

The total number of parameters in the M4 process in (4.1) is 18. There is a nuisance parameter that is the variance of N_{id}. However, we do not need to estimate the nuisance parameter in order to estimate the values of M4 model parameters. We first generate data by simulating these bivariate processes, then based on the simulated data we re-estimate all coefficients simultaneously and compute their asymptotic covariance matrix. Table 1 is obtained using simulated data with a sample size of 10,000. The x_{jd} and $x'_{j'd}$ values are determined using the procedure described earlier, and subsequently, values of $\widehat{b}_d(x_{jd})$, and $\widehat{b}_{1d'}(x'_{j'd})$ can be estimated.

Now we have a system of nonlinear equations, whose variables are the values of $\widehat{a}_{l,k,d}$. The estimates are found through a Monte Carlo optimization algorithm. For all l, k, d, we simulate 5000 vector values of $\widehat{a}_{l,k,d}$s from which the ratios $\widehat{a}_{l,k+1,d}/\widehat{a}_{l,k,d}$ and $\widehat{a}_{l,k,1}/\widehat{a}_{l,k,d}$ $(d > 1)$, are falling in the regions determined by x_{jd}, $x'_{j'd}$ computed in Steps 3 and 4. We keep the vector whose ratios $\widehat{a}_{l,k+1,d}/\widehat{a}_{l,k,d}$, $\widehat{a}_{l,k,1}/\widehat{a}_{l,k,d}$ have the minimal distance to the averaged ratios (obtained in Steps 3 and 4). We repeat this process 100 times, and hence we get 100 vectors. We keep the vector which gives the minimal distance between theoretical values of $b_d(x)$ and $b_{1d}(x)$ computed using the kept $\widehat{a}_{l,k,d}$ values to the estimated functions $\widehat{b}_d(x)$ and $\widehat{b}_{1d}(x)$.

In Table 1, the estimated values of almost all cases (except $a_{1,-1,1}$, $a_{1,0,1}$) are very close to the true parameter values. The estimated values in cases of $a_{1,-1,1}$, $a_{1,0,1}$ were probably affected by the added noises N_{id} and clustering methods. Apart from that, the estimates are close to the true values. We believe these results demonstrate the efficacy of the proposed estimating procedures.

To examine how sensitive this procedure is to the noise standard deviation σ, the simulation has been repeated for values of σ from 0.1 to 3.5 in increments of 0.1. The estimation procedure is reasonably stable up to $\sigma = 2$, though the parameter SEs increase with σ. For $\sigma > 2$, the optimization routines failed, mainly because of the initial clustering procedure. Figure 3 illustrates the increase of the maximum simulated standard error across all parameters as a function of σ.

5 Modeling jumps in returns of financial assets

We consider a trivariate time series of exchange returns from the Japanese YEN against the US dollar (JPY/USD), the Canadian dollar against the US dollar (CAD/USD), and the British pound against the US dollar (GBP/USD). They are plotted in Figure 4.

Parameter	True value	Estimated value	Standard deviation	Parameter	True value	Estimated value	Standard deviation
$a_{1,-1,1}$	.1500	0.0941	0.0418	$a_{1,-1,2}$	.0700	0.0379	0.0295
$a_{1,0,1}$	.2000	0.1258	0.0351	$a_{1,0,2}$	.0400	0.0231	0.0108
$a_{1,1,1}$	.0200	0.0101	0.0486	$a_{1,1,2}$	.0300	0.0166	0.0137
$a_{2,-1,1}$	.0500	0.0573	0.0573	$a_{2,-1,2}$	.1000	0.0949	0.0215
$a_{2,0,1}$	.1000	0.1132	0.0558	$a_{2,0,2}$	.1300	0.1387	0.0417
$a_{2,1,1}$	.0300	0.0295	0.0456	$a_{2,1,2}$	.1700	0.1765	0.0518
$a_{3,-1,1}$	.1600	0.2091	0.0459	$a_{3,-1,2}$	.1100	0.1272	0.0323
$a_{3,0,1}$	.1700	0.2143	0.0610	$a_{3,0,2}$	.1200	0.1325	0.0318
$a_{3,1,1}$	.1200	0.1468	0.0582	$a_{3,1,2}$	.2300	0.2527	0.0707

Table 1: *Simulation results for model (4.1).* x_{j1} *=(0.0775, 0.1550, 0.2582, 0.4120, 0.6163, 0.8058, 0.9806, 1.1363, 1.2728, 1.5032, 1.8274 , 2.1884).* x_{j2} *=(0.4421, 0.6287, 0.7071, 0.8318, 1.0027, 1.1423, 1.2505, 1.3050, 1.3057, 1.4641, 1.7802, 2.1321).* $x'_{j'}$ *=(0.1376, 0.2642, 0.4258, 0.5106, 0.5189, 0.5631, 0.6433, 0.7055, 0.7496, 0.9356, 1.2637, 1.4312, 1.4383, 1.6236, 1.9871, 2.8617, 4.2474, 5.4343). Standard deviations are computed by applying Theorem 3.3.*

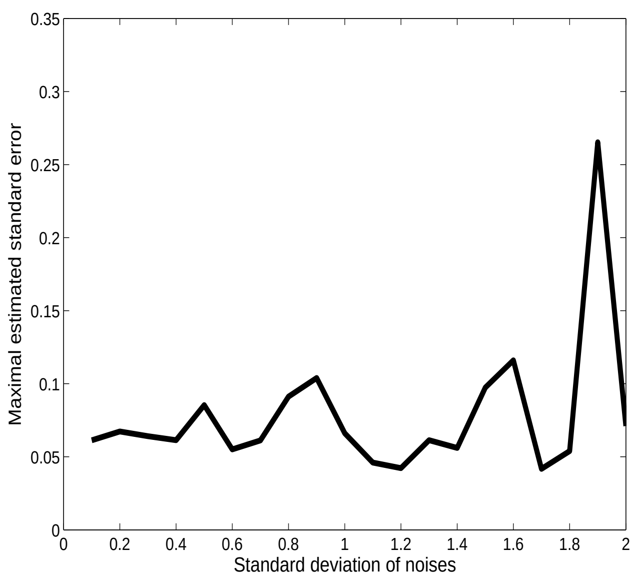

Figure 3: Maximal simulated standard error of parameters versus standard deviations of N_{id}.

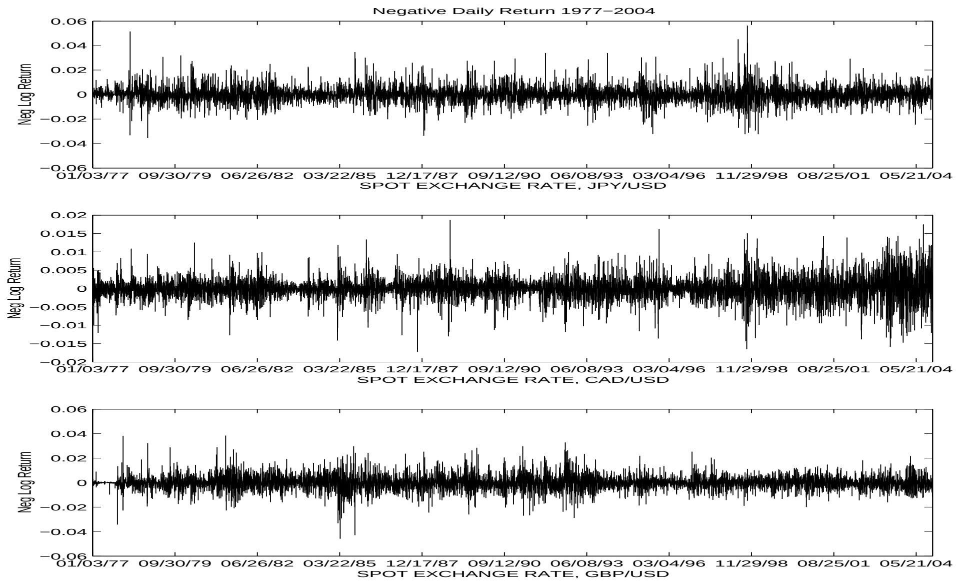

Figure 4: *Negative daily returns. The top plot is for JPY/USD, the middle plot is for CAD/USD, and the bottom plot is for GBP/USD.*

There are extremal observed values in each series but there are also changes in volatility. As a first step in the analysis, we propose a procedure to remove the volatility.

5.1 Data transformation

To remove volatility, we propose a simple application of the GARCH (generalized autoregressive conditional heteroscedasticity) model, originally proposed by Bollerslev (1986). Lee and Hansen (1994) discussed maximum likelihood estimation of GARCH model with weak stationary residuals; Mikosch (2003) gave a very thorough study of GARCH modeling of dependence and tails of financial time series. Cross-sectional dependencies between GARCH residuals have been studied by McNeil and Frey (2000) and Engle (2002). For the present analysis, we are not assuming that the series are GARCH, but we use a GARCH(1,1) model as a tool to model volatilities. The estimated conditional standard deviations are shown in Figure 5. The original data sets are then divided by these standard deviations and three new standardized time series – GARCH residuals – are obtained. The standardized time series (not shown) appear stationary.

The next step is to transform the series to have unit Fréchet marginal distributions. Smith (1989) showed how a process of exceedances over a high threshold can be modeled in terms of the limiting generalized extreme value (GEV) distribution function of form

$$H(x) = \exp\left\{ -\left(1 + \xi \frac{x-\mu}{\psi}\right)_+^{-1/\xi} \right\}, \tag{5.1}$$

where μ is a location parameter, $\psi > 0$ is a scale parameter, and ξ is a shape parameter.

This model is used to fit the data above a certain threshold (.02 for the original data, 1.2 for the volatility-standardized data) for each series. Standard diagnostics such as those in Smith and Shively (1995), Tsay (1999) and Smith (2003) show a poor fit to the extreme value model based on

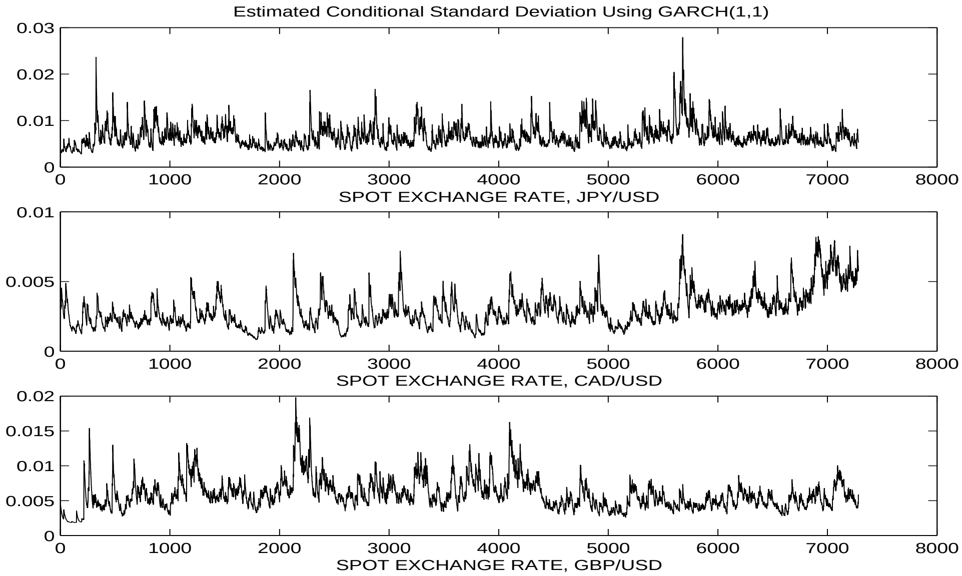

Figure 5: *Estimated volatility by GARCH. The top plot is for JPY/USD, the middle plot is for CAD/USD, and the bottom plot is for GBP/USD.*

the original data, but a much better fit using the standardized data formed from residuals of the GARCH(1,1) process.

In the finance literature, positive returns and negative returns are often tested to be asymmetric. We first fitted the standardized positive returns and negative returns to GEV separately, and found that all fitted return distributions are not significantly different from Gumbel type distributions, and positive returns and negative returns can be considered approximately symmetric. For this reason, we fit the standardized absolute returns to GEV. The estimated parameter values of the GEV distributions are summarized in Table 2.

After completing these transformations we have a devolatized time series of absolute returns standardized to unit Fréchet margins. The next stage of the analysis will fit an M4 model to this transformed process.

Series	N_u	μ (SE)	$\log\psi$ (SE)	ξ (SE)
JPY/USD	542	0.396463	-0.733705	0.111582
		(0.139571)	(0.160542)	(0.047574)
CAD/USD	539	0.501447	-0.910765	0.142192
		(0.127555)	(0.170876)	(0.050466)
GBP/USD	546	0.352415	-0.710196	0.104102
		(0.123285)	(0.137673)	(0.040217)

Table 2: *Estimation of parameters in GEV using standardized absolute return series. The notation N_u means the number of observations over the threshold u where 1.69 for JPY/USD, 1.63 for CAD/USD, 1.66 for GBP/USD.*

5.2 Model selection

Key parameters for determining the model are L_d, the number of clusters in the time series for component d, and the range parameters K_{1ld} and K_{2ld}. In practice, rather than seeking optimal estimates of these parameters, we advocate examining the data for evidence of clustering in exceedances over a high threshold, then choosing a model that is consistent with the pattern of observed clusters. We illustrate these ideas based on the transformed JPY/USD data Y_{i1}, CAD/USD data Y_{i2}, and GBP/USD data Y_{i3}, $i = 1, 2, \ldots, 5715$ (the total number of days on which the prices change).

The threshold for determining the significance of tail dependence is not necessarily the same as the threshold used for transforming the marginal distributions. As an example, Figure 6 shows empirical tail dependence indexes (see Section 2) for pairs of series on the same day, computed for a range of thresholds from 18.5 to 21 (or 94.7 percentile to 95.3 percentile of unit Fréchet distribution respectively). One can see in this range, the empirical tail dependence indexes suggest variable tail dependence. Based on this, we select $u = 19.5$ (or 95th percentile) for subsequent analysis. For that threshold, we find that the maximal range of consecutive days for which the jumps in returns are over the threshold value are 2, 2, 3 days (Columns 2, 4 in Table 3) for JPY/USD, CAD/USD, and GBP/USD respectively.

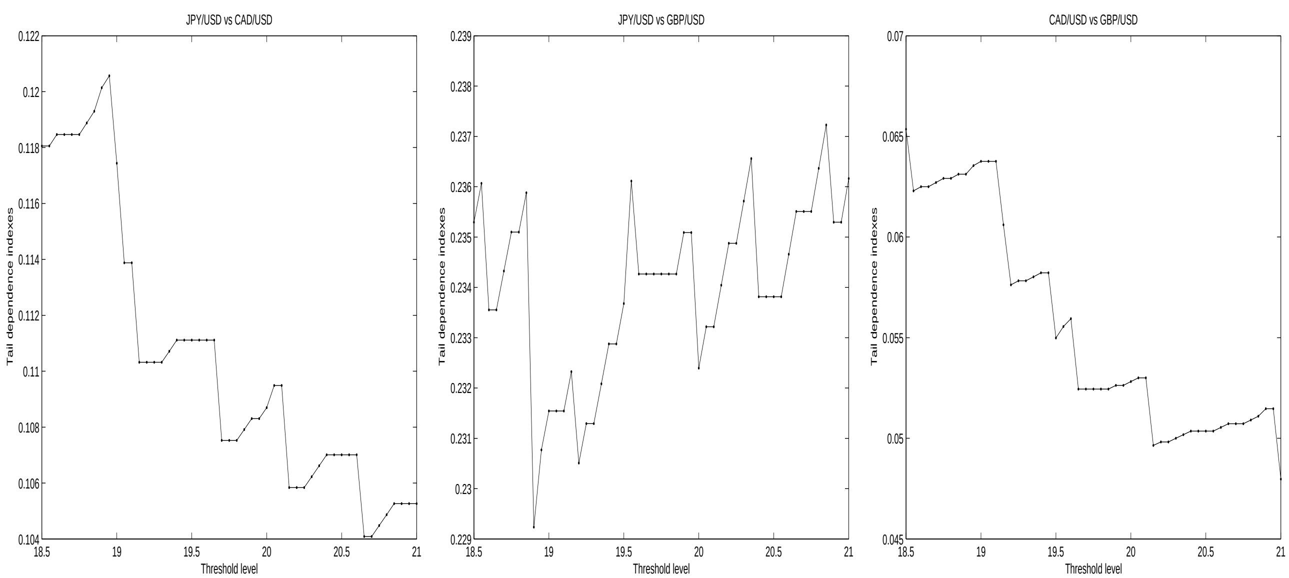

Figure 6: *Empirical estimates of tail dependence indexes against threshold values. The left panel (respectively, the middle panel, and the right panel) is for JPY/USD return and CAD/USD return (respectively, JPY/USD and GBP/USD, and CAD/USD and GBP/USD) at the same day.*

A more detailed table of joint exceedances is in Table 3. This can be used to examine which dependencies are statistically significant. For example, the count of $\{Y_{i1} > u, Y_{i2} > u\}$ is 31. A 2×2 table classifying all pairs (Y_{i1}, Y_{i2}) according to whether they are below or above the threshold, has entries 5203, 248, 233, 31. Fisher's exact test of independence, for this table, has a p-value of about 5×10^{-6}. Based on this, we conclude that the dependence between the events $Y_{i1} > u$ and $Y_{i2} > u$ is real. The same conclusion holds for the dependences $Y_{i1} > u$ and $Y_{i3} > u$, and for $Y_{i2} > u$ and $Y_{i+1,2} > u$, but not the other pairwise dependences in Table 3. Nevertheless, in considering serial

$\{Y_{i1} > u, Y_{i2} > u\}$ **31**	$\{Y_{i1} > u, Y_{i3} > u\}$ **68**	$\{Y_{i2} > u, Y_{i3} > u\}$ **16**	$\{Y_{i1} > u, Y_{i2} > u, Y_{i3} > u\}$ **6**
$\{Y_{i1} > u\}$ **279**	$\{Y_{i1} > u, Y_{i+1,1} > u\}$ **12**		$\{Y_{i1} > u, Y_{i+1,1} > u, Y_{i+2,1} > u\}$ **0**
$\{Y_{i2} > u\}$ **264**	$\{Y_{i2} > u, Y_{i+1,2} > u\}$ **20**		$\{Y_{i2} > u, Y_{i+1,2} > u, Y_{i+2,2} > u\}$ **0**
$\{Y_{i3} > u\}$ **291**	$\{Y_{i3} > u, Y_{i+1,3} > u\}$ **13**		$\{Y_{i3} > u, Y_{i+1,3} > u, Y_{i+2,3} > u\}$ **2**

Table 3: *Counts of days that the absolute returns are over a threshold value in consecutive days. All counts are mutually exclusive.*

dependence within each series, there is no obvious reason why the CAD/USD series should behave differently from the other two, so in the subsequent discussion we assume lag-1 serial dependence within each series. We also look at triple exceedances $\{Y_{i1} > u, Y_{i2} > u, Y_{i3} > u\}$ — in this case the observed number (6) is also statistically significant based on the null hypothesis of independence (given the observed frequency of exceedances of the threshold by each of the marginal series, the expected count of triple exceedances should be about 0.75 under the null hypothesis of independence; for a Poisson variable of mean 0.75, the probability of observing a value 6 or larger is about 0.0001). We also checked pairs $\{Y_{id},\ Y_{i+j,d}\}$ for $j > 1$ and found the number of joint exceedances do not suggest evidence of dependence. Based on these considerations, we propose a model with $L = 10$ clusters and the configuration of non-zero coefficients given in Table 4.

The estimates are derived using the methods described in Sections 3 and 4. However, this is a case where Theorem 3.3 is not directly applicable and we use Corollary 3.4 instead. If signature pattern 7 in Table 4 were not present, we would be able to use Theorem 3.3 directly, but this does not seem a realistic model given the evidence that simultaneous exceedances of the threshold by all three series occur at a greater rate than would be observed by chance.

If we only consider a univariate process, the number of signature patterns is 2 with the first signature pattern being a 2-day dependence pattern, and the second signature pattern consisting of a single exceedance of the threshold. Also, Remark 1 determines the relative frequency of the two signature patterns for a univariate process. Using the $D = 1$ case of Theorem 3.3, therefore, we estimate parameters $a^*_{1,-1,d}$, $a^*_{1,0,d}$, $a^*_{2,0,d}$, for each of $d = 1, 2, 3$. For example, in the case $d = 1$ we have $\widehat{a}^*_{1,-1,1} = 0.0165$, $\widehat{a}^*_{1,0,1} = 0.0122$, $\widehat{a}^*_{2,0,1} = 0.9713$. The full set of parameters $\{a_{l,k,d}\}$ is estimated as described in the discussion preceding Corollary 3.4. For example, we approximate $a_{7,k,1} = \frac{6}{279} a^*_{2,k,1}$ on the basis that, out of all single exceedances of the threshold by component 1, a fraction $\frac{6}{279}$ are triple exceedances by all three components. Thus in this case we write $l^* = 2$ and $\beta_{l,1} = \frac{6}{279}$. A more thorough error analysis would also take into account that $\frac{6}{279}$ is itself an estimate but we do not do that here as our main intention is to illustrate the application of the asymptotic formulae of Section 3. To take another example, we approximate $a_{5,k,1} = \frac{62}{279} a^*_{2,k,1}$ on the basis that in counting instances of signature pattern 5, we do not count the overlap with signature pattern 7. Thus $\beta_{5,1} = \frac{62}{279}$. By proceeding through all 3 components and all 10 signature patterns in similar fashion, the full Table 4 is constructed.

Table 4: *Estimation of parameters in M4 model applied to standardized exchange rate time series. The values in parentheses are standard errors.*

Signature	JPY/USD		CAD/USD		GBP/USD	
l	$a_{l,-1,1}$	$a_{l,0,1}$	$a_{l,-1,2}$	$a_{l,0,2}$	$a_{l,-1,3}$	$a_{l,0,3}$
1	0.0165	0.0122				
	(0.1212)	(0.0251)				
2			0.0304	0.0075		
			(0.1425)	(0.0278)		
3					0.0311	0.0170
					(0.0398)	(0.0267)
4		0.0870		0.0911		
		(0.0103)		(0.0127)		
5		0.2159				0.2028
		(0.0256)				(0.0104)
6				0.0364		0.0327
				(0.0051)		(0.0017)
7		0.0209		0.0219		0.0196
		(0.0025)		(0.0031)		(0.0010)
8		0.6476				
		(0.0768)				
9				0.8127		
				(0.1136)		
10						0.7196
						(0.0356)

5.3 A new co-movement measure and its estimation

As an illustration of how the methods of this paper might be used to calculate quantities of practical interest, we consider the following extreme co-movement measure:

$$\lambda(t,T) = \lim_{\mathbf{u}\nearrow\mathbf{x}_F} \Pr\{\xi(t,T,\mathbf{u}) \geq 2 | \xi(0,t,\mathbf{u}) \geq 1\} \tag{5.2}$$

where $\mathbf{x}_F$ is the right end point of the distribution function F and

$$\xi(t,T,\mathbf{u}) = \max_{t\leq i\leq T} \sum_{d=1}^{D} I_{(Y_{id}>u_d)}. \tag{5.3}$$

Thus the idea is to estimate the maximum number of joint exceedances in the time period t to T given at least one exceedance in $(0,t)$. The case $t=T=0$ and $D=2$ is the usual tail dependence function in the literature (Embrechts, Lindskog, and McNeil 2001).

An obvious approach to estimating $\lambda(t,T)$ is simply to pick a vector of thresholds $\mathbf{u}$ and estimate (5.2) empirically, by counting exceedances. However, this will not work if $\mathbf{u}$ is too high. An alternative "model-based" approach is first to fit an M4 process, and then estimate (5.2) from simulations of the process. In this case, there is no theoretical limit on $\mathbf{u}$, because we may keep simulating the process until there are sufficient exceedances of the threshold.

In Table 5, we compare the empirical and model-based estimates for thresholds determined as the 90th and 97.5th percentiles of each component, and show the model-based results for the 99.5th percentile which is beyond the range of which the empirical estimate may be computed.

Table 5: *Computed empirical values of extreme co-movement measure in (5.2), where* $t = 1$.

$T - t =$	1	2	3	4	5	6	Remark
Data	0.8848	0.9480	0.9535	0.9535	0.9535	0.9535	90th percentile
	(.056)	(.037)	(.032)	(.032)	(.032)	(.032)	
Model	0.9088	0.9446	0.9455	0.9455	0.9455	0.9455	
	(.008)	(.006)	(.006)	(.006)	(.006)	(.006)	
Data	0.8984	0.9805	0.9922	0.9922	0.9922	0.9922	97.5th percentile
	(.088)	(.048)	(.031)	(.031)	(.031)	(.031)	
Model	0.9425	0.9801	0.9811	0.9811	0.9811	0.9811	
	(.014)	(.008)	(.008)	(.008)	(.008)	(.008)	
Model	0.9629	0.9939	0.9956	0.9956	0.9956	0.9956	99.5th percentile
	(.066)	(.025)	(.019)	(.019)	(.019)	(.019)	

From Table 5, we see that each row has an increasing order and after 3 days the measures remain the same. This suggests that after a particular time, the price history does not provide further useful information for extreme price movements. Also considering the estimated SEs (in parentheses), the model based measures are good approximations to the data based measures. This suggests that M4 models and our proposed statistical inference methods could be used in risk management where a large price change in any one of several assets could be of critical importance in managing a portfolio.

6 Discussion

The main contribution of the paper is to propose the use of M4 process for modeling joint extremal behavior in financial time series that show dependence both between the series and in time. Because there are many parameters in the model and maximum likelihood methods are not applicable, identifiability of the model in terms of its bivariate distributions plays a major role in constructing estimators based on empirical functions. We have given sufficient conditions (Proposition 2.2) using functions $b_{dd'}(x)$ for $d = 1$ and d' variable — more generally, we might need to use the full class of $\{b_{dd'}\}$ functions, but Proposition 2.2 is easily extended to that case. Our main estimation results are in Theorem 3.3 and Corollary 3.4, where the latter is based on a slightly simplified estimation procedure that avoids some of the identifiability issues. The mathematical theory leaves open the optimal determination of the various model order parameters and the x_{jd}, $x'_{j'd'}$ used in Theorem 3.3, but we have tried to show in Sections 4 and 5 how these may be determined in practice.

There are many possible applications to risk management or extensions of the concept of value at risk to multiple time series. In Section 5.3, we illustrated this with one possible measure of extreme co-movements, but there are many other measures of interest for which similar statistical methods could be applied.

7 Appendix of notations used in Section 3

Notations and their expressions in this section are used to establish Lemma 3.1. These notations and their expressions are:

$$\begin{aligned}\mu_{djd} = E\big[U_d(x_{jd})\big] &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}),\\ &\quad d = 1, \ldots, D,\ j = 1, \ldots, m,\end{aligned}$$

$$\begin{aligned}\mu_{1d'j'd'} = E\big[U_{1d'}(x'_{j'd'})\big] &= \Pr(Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}),\\ &\quad d' = 2, \ldots, D,\ j' = 1, \ldots, m',\end{aligned}$$

$$\begin{aligned}\mu_{djd,\ d'j'd'} &= E\Big[\big(I_{\{Y_{1d}\le 1, Y_{2d}\le x_{jd}\}} - \mu_{djd}\big)\big(I_{\{Y_{1d'}\le 1, Y_{2d'}\le x_{j'd'}\}} - \mu_{d'j'd'}\big)\Big]\\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{1d'} \le 1,\ Y_{2d'} \le x_{j'd'}) - \mu_{djd}\mu_{d'j'd'},\\ &\quad d, d' = 1, \ldots, D,\ j, j' = 1, \ldots, m,\end{aligned}$$

$$\begin{aligned}\mu_{djd,\ 1d'j'd'} &= E\Big[\big(I_{\{Y_{1d}\le 1, Y_{2d}\le x_{jd}\}} - \mu_{djd}\big)\big(I_{\{Y_{11}\le 1, Y_{1d'}\le x'_{j'd'}\}} - \mu_{1d'j'd'}\big)\Big]\\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}) - \mu_{djd}\mu_{1d'j'd'},\\ &\quad d = 1, \ldots, D,\ j = 1, \ldots, m,\\ &\quad d' = 2, \ldots, D,\ j' = 1, \ldots, m',\end{aligned}$$

$$\begin{aligned}\mu_{1d'j'd',\ djd} &= E\Big[\big(I_{\{Y_{11}\le 1, Y_{1d'}\le x'_{j'd'}\}} - \mu_{1d'j'd'}\big)\big(I_{\{Y_{1d}\le 1, Y_{2d}\le x_{jd}\}} - \mu_{djd}\big)\Big]\\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{11} \le 1,\ Y_{1d'} \le x'_{j'd'}) - \mu_{djd}\mu_{1d'j'd'}\\ &= \mu_{djd,\ 1d'j'd'},\\ &\quad d = 1, \ldots, D,\ j = 1, \ldots, m,\\ &\quad d' = 2, \ldots, D,\ j' = 1, \ldots, m',\end{aligned}$$

$$\begin{aligned}\mu_{1djd,\ 1d'j'd'} &= E\Big[\big(I_{\{Y_{11}\le 1, Y_{1d}\le x'_{jd}\}} - \mu_{1djd}\big)\big(I_{\{Y_{11}\le 1, Y_{1d'}\le x'_{j'd'}\}} - \mu_{1d'j'd'}\big)\Big]\\ &= \Pr(Y_{11} \le 1,\ Y_{1d} \le x'_{jd},\ Y_{1d'} \le x'_{j'd'}) - \mu_{1djd}\mu_{1d'j'd'},\\ &\quad d = 2, \ldots, D,\ j = 1, \ldots, m',\\ &\quad d' = 2, \ldots, D,\ j' = 1, \ldots, m',\end{aligned}$$

$$\begin{aligned}w^{(k)}_{djd,\ d'j'd'} &= E\Big[\big(I_{\{Y_{1d}\le 1, Y_{2d}\le x_{jd}\}} - \mu_{djd}\big)\big(I_{\{Y_{1+k,d'}\le 1, Y_{2+k,d'}\le x_{j'd'}\}} - \mu_{d'j'd'}\big)\Big]\\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{1+k,d'} \le 1,\ Y_{2+k,d'} \le x_{j'd'}) - \mu_{djd}\mu_{d'j'd'},\\ &\quad d, d' = 1, \ldots, D,\ j, j' = 1, \ldots, m,\end{aligned}$$

$$\begin{aligned}w^{(k)}_{djd,\ 1d'j'd'} &= E\Big[\big(I_{\{Y_{1d}\le 1, Y_{2d}\le x_{jd}\}} - \mu_{djd}\big)\big(I_{\{Y_{1+k,1}\le 1, Y_{1+k,d'}\le x'_{j'd'}\}} - \mu_{1d'j'd'}\big)\Big]\\ &= \Pr(Y_{1d} \le 1,\ Y_{2d} \le x_{jd}, Y_{1+k,1} \le 1,\ Y_{1+k,d'} \le x'_{j'd'}) - \mu_{djd}\mu_{1d'j'd'},\\ &\quad d = 1, \ldots, D,\ j = 1, \ldots, m,\\ &\quad d' = 2, \ldots, D,\ j' = 1, \ldots, m',\end{aligned}$$

$$
\begin{aligned}
w^{(k)}_{1d'j'd',\ djd} &= E\Big[\big(I_{\{Y_{1,1}\le 1, Y_{1,d'}\le x'_{j'd'}\}} - \mu_{1d'j'd'}\big)\big(I_{\{Y_{1+k,d}\le 1, Y_{2+k,d}\le x_{jd}\}} - \mu_{djd}\big)\Big] \\
&= \Pr(Y_{1,1}\le 1,\ Y_{1,d'}\le x'_{j'd'},\ Y_{1+k,d}\le 1,\ Y_{2+k,d}\le x_{jd}) - \mu_{djd}\mu_{1d'j'd'}, \\
&\quad\ d=1,\dots,D,\ j=1,\dots,m, \\
&\quad\ d'=2,\dots,D,\ j'=1,\dots,m',
\end{aligned}
$$

$$
\begin{aligned}
w^{(k)}_{1djd,\ 1d'j'd'} &= E\Big[\big(I_{\{Y_{11}\le 1, Y_{1d}\le x'_{jd}\}} - \mu_{1djd}\big)\big(I_{\{Y_{1+k,1}\le 1, Y_{1+k,d'}\le x'_{j'd'}\}} - \mu_{1d'j'd'}\big)\Big] \\
&= \Pr(Y_{11}\le 1,\ Y_{1d}\le x'_{jd}, Y_{1+k,1}\le 1,\ Y_{1+k,d'}\le x'_{j'd'}) - \mu_{1djd}\mu_{1d'j'd'}, \\
&\quad\ d=2,\dots,D,\ j=1,\dots,m', \\
&\quad\ d'=2,\dots,D,\ j'=1,\dots,m'.
\end{aligned}
$$

Based on the above quantities we define the following vectors:

$$
\boldsymbol{\mu} = \begin{bmatrix} \mu_{111} \\ \mu_{121} \\ \vdots \\ \mu_{1m1} \\ \mu_{212} \\ \vdots \\ \mu_{DmD} \\ \mu_{1212} \\ \mu_{1222} \\ \vdots \\ \mu_{12m'2} \\ \mu_{1313} \\ \vdots \\ \mu_{1Dm'D} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \\ \mu_{m+1} \\ \vdots \\ \mu_{D\times m} \\ \mu_{D\times m+1} \\ \mu_{D\times m+2} \\ \vdots \\ \mu_{D\times m+m'} \\ \mu_{D\times m+m'+1} \\ \vdots \\ \mu_{D\times m+(D-1)m'} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1(x_{11}) \\ b_1(x_{21}) \\ \vdots \\ b_1(x_{m1}) \\ b_2(x_{12}) \\ \vdots \\ b_D(x_{mD}) \\ b_{12}(x'_{12}) \\ b_{12}(x'_{22}) \\ \vdots \\ b_{12}(x'_{m'2}) \\ b_{13}(x'_{13}) \\ \vdots \\ b_{1D}(x'_{m'D}) \end{bmatrix}.
$$

Notice that the subscripts of the elements of vector $\boldsymbol{\mu}$ are different in its two vector forms though the same notation μ is used. To form the above vectors, for example, to get the rth value in the vector $\boldsymbol{\mu}$, we have used the following index transformation:

$$
\begin{cases}
\mu_{djd} \ \rightarrow\ \mu_r, & \text{where } r = (d-1)\times m + j, \\
\mu_{1d'jd'} \ \rightarrow\ \mu_r, & \text{where } r = D\times m + (d'-2)\times m' + j,
\end{cases}
$$

where [.] takes integer values. We now use the similar relations between the indexes of μ_{djd} and the indexes of μ_r to define the following variables:

$$
\sigma_{rs} = \begin{cases}
\mu_{djd,d'j'd'}, & \text{if } r\le D\times m, s\le D\times m, \\
\mu_{djd,1d'j'd'}, & \text{if } r\le D\times m, s> D\times m, \\
\mu_{1djd,d'j'd'}, & \text{if } r> D\times m, s\le D\times m, \\
\mu_{1djd,1d'j'd'}, & \text{if } r> D\times m, s> D\times m,
\end{cases}
$$

$$
w_k^{rs} = \begin{cases} w^{(k)}_{djd,d'j'd'}, & \text{if } r \le D\times m, s \le D\times m, \\ w^{(k)}_{djd,1d'j'd'}, & \text{if } r \le D\times m, s > D\times m, \\ w^{(k)}_{1djd,d'j'd'}, & \text{if } r > D\times m, s \le D\times m, \\ w^{(k)}_{1djd,1d'j'd'}, & \text{if } r > D\times m, s > D\times m, \end{cases}
$$

and the matrices:

$$
\Sigma = (\sigma_{rs}),\ W_k = (w_k^{rs}),\ \Theta = (\mathrm{diag}\{\mu\})^{-1} \times (\mathrm{diag}\{\mathbf{x}\}).
$$

8 Technical arguments

Proof of Proposition 2.2. We first prove $q_d(x)$ (or equivalently $b_d(x)$) uniquely determine all coefficients in dth process. Since all the ratios are different and are points at which $q_d(x)$ changes slopes or $q_d'(x)$ has jumps. So based on the jump points of $q_d(x)$, the ratios of $\frac{a_{l,j+1,d}}{a_{l,j,d}}$ are uniquely determined. Let's now rewrite $q_d(x)$ as

$$
q_d(x) = \sum_{l=1}^{L_d} x b_{ld} \sum_{m=1-K_{2ld}}^{2+K_{1ld}} \max\big(c_{l,1-m,d}, \frac{c_{l,2-m,d}}{x}\big). \tag{8.1}
$$

where $\sum_j c_{l,j,d} = 1$ for each l and all $c_{l,j,d}$ are uniquely determined by the ratios which are the slope change points of $q_d(x)$.

Suppose now $q_d(x)$ has a different representation, say

$$
q_d(x) = \sum_{l=1}^{L_d} x b'_{ld} \sum_{m=1-K_{2ld}}^{2+K_{1ld}} \max\big(c_{l,1-m,d}, \frac{c_{l,2-m,d}}{x}\big) \tag{8.2}
$$

then

$$
\sum_{l=1}^{L_d} (b_{ld} - b'_{ld}) \sum_{m=1-K_{2ld}}^{2+K_{1ld}} \max\big(c_{l,1-m,d}, \frac{c_{l,2-m,d}}{x}\big) = 0 \tag{8.3}
$$

for all x.

Suppose we have chosen $x_1,\ x_2, \dots, x_{L_d-1}$ and formed the matrix

$$
\Delta_d = \begin{bmatrix} \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_1}) & \cdots & \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{L_d,1-m,d}, \frac{c_{L_d,2-m,d}}{x_1}) \\ \vdots & \ddots & \vdots \\ \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_{L-1}}) & \cdots & \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{L_d,1-m,d}, \frac{c_{L_d,2-m,d}}{x_{L-1}}) \\ 1 & \cdots & 1 \end{bmatrix}
$$

and set

$$
\begin{bmatrix} \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_1}) & \cdots & \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{L_d,1-m,d}, \frac{c_{L_d,2-m,d}}{x_1}) \\ \vdots & \ddots & \vdots \\ \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_{L-1}}) & \cdots & \sum\limits_{m=1-K_{2ld}}^{2+K_{1ld}} \max(c_{L_d,1-m,d}, \frac{c_{L_d,2-m,d}}{x_{L-1}}) \\ 1 & \cdots & 1 \end{bmatrix} \begin{pmatrix} b_{1d} - b'_{1d} \\ \vdots \\ \vdots \\ b_{L_d d} - b'_{L_d d} \end{pmatrix} = 0.
$$

$|\Delta_d|$ is the determinant of the system of linear equations. Assume now the L_d determinants of the $(L_d-1)\times(L_d-1)$ matrices formed from the bottom L_d-1 rows are not all zero. Since $c_{l,k,d}$ are known and $\sum_{m=1-K_{2ld}}^{2+K_{1ld}} c_{l,i-m,d} = 1,\ i = 1, 2$, then there exist $x_{\min}$ and $x_{\max}$ such that when $x_1 < x_{\min}$ or $x_1 > x_{\max}$, all elements of first row in Δ_d are $\frac{1}{x_1}$ or 1 respectively. This will give two constant rows in $|\Delta_d|$, so when $x_1 < x_{\min}$ or $x_1 > x_{\max}$, we have $|\Delta_d| = 0$. When x_1 varies in $[x_{\min}, x_{\max}]$, denoting Δ_d by $\Delta_d(x_1)$, then

$$|\Delta_d(x_1)| = \frac{1}{x_1}\sum c_{i,j,d}|\Delta_d|_{1j} + \sum c_{i',j',d}|\Delta_d|_{1j'} \tag{8.4}$$

where $|\Delta_d|_{1j} \neq 0$, $|\Delta_d|_{1j'} \neq 0$ are the $(1, j)$ or $(1, j')$ minors of Δ_d. Both summations in the right hand side of (8.4) are over all non-zero minors of the first row of Δ_d and the corresponding $\frac{c_{i,j,d}}{x_1}$ or $c_{i',j',d}$. If $|\Delta_d(x_1)| = 0$, by varying x_1 in $[x_{\min}, x_{\max}]$, at some point x, some $\frac{1}{x_1}c_{i,j,d}|\Delta_d|_{1j}$ of the summation $\frac{1}{x_1}\sum c_{i,j,d}|\Delta_d|_{1j}$ change to $c_{i',j',d}|\Delta_d|_{1j'}$ and add to $\sum c_{i',j',d}|\Delta_d|_{1j'}$, or *vice versa*, and this change results in $|\Delta_d(x)| \neq 0$. Hence it cannot be true that $|\Delta_d| = 0$ for all x_1. This argument can be applied to lower dimension matrices. On the other hand, we can start from a 2×2 matrix and extend it to $L_d \times L_d$ matrix such that the determinant is not zero as required. Therefore, there exist constants $x_1, \ldots, x_{L_d-1}$ such that each system of linear equations has a unique solution. We then conclude $b_{ld} = b'_{ld}$, for all l. So $q_d(x)$ uniquely determine all $a_{l,j,d}$.

Now, we prove the results for bivariate maxima and moving maxima processes. Since $b_d(x)$ and $b_{d'}(x)$ uniquely determine all values of parameters $a_{l,k,d}$ and $a_{l,k,d'}$ respectively, we can get

$$(a_{l,-K_{1ld},d},\ a_{l,-K_{1ld}+1,d}, \ldots, a_{l,K_{2ld},d}),\ l = 1, \ldots, L_d$$

and

$$(a_{l',-K_{1ld},d'},\ a_{l',-K_{1ld}+1,d'}, \ldots, a_{l',K_{2ld},d'}),\ l' = 1, \ldots, L_d.$$

Since all nonzero existing ratios $\frac{a_{l,k,d}}{a_{l,k,d'}}$ are distinct, any permutation of index l or index k in the triple subindex of $a_{l,k,d'}$ will result in different ratios which will be different from the jump points of $q_{dd'}(x)$, so the jump points of $q_{dd'}(x)$ uniquely determine

$$\Big(\frac{a_{l,-K_{1ld},d}}{a_{l,-K_{1ld},d'}},\ \frac{a_{l,-K_{1ld}+1,d}}{a_{l,-K_{1ld}+1,d'}},\ \cdots,\ \frac{a_{l,K_{2ld},d}}{a_{l,K_{2ld},d'}}\Big)$$

for all l. So (2.6) and (2.7) eventually uniquely determine all the true values of all parameters $a_{l,k,d}$ and $a_{l,k,d'}$.

The reason why $x_1, x_2, \ldots, x_m$ uniquely determine all values of $a_{l,k,d}$ and $a_{l,k,d'}$ is because $q_d(x)$, $q_{d'}(x)$ and $q_{dd'}(x)$ are piecewise linear functions which can be uniquely determined by a finite number of points as long as there are at least two points between any two jump points.

Using the same arguments above, we can prove the results for $D > 2$. □

Proof of Lemma 3.1. We use Theorem 27.4 in Billingsley (1995), re-stated below as Results **TA1**.

For a sequence $\zeta_1, \zeta_2, \ldots$ of random variables, let α_n be a number such that

$$|\Pr(A \cap B) - \Pr(A)\Pr(B)| \leq \alpha_n$$

for $A \in \sigma(\zeta_1, \ldots, \zeta_k)$, $B \in \sigma(\zeta_{k+n}, \zeta_{k+n+1}, \ldots)$, and $k \geq 1, n \geq 1$. When $\alpha_n \to 0$, the sequence $\{\zeta_n\}$ is said to be α-mixing. This means that ζ_k and ζ_{k+n} are approximately independent for large n.

Results **TA1** (Theorem 27.4 in Billingsley (1995)). Suppose that $\Upsilon_1, \Upsilon_2, \ldots,$ is stationary and α-mixing with $\alpha_n = O(n^{-5})$ and that $E[\Upsilon_n] = 0$ and $E[\Upsilon_n^{12}] < \infty$. If $S_n = \Upsilon_1 + \cdots + \Upsilon_n$, then

$$n^{-1}\text{Var}[S_n] \to \sigma^2 = E[\Upsilon_1^2] + 2\sum_{k=1}^{\infty} E[\Upsilon_1 \Upsilon_{1+k}],$$

where the series converges absolutely. If $\sigma > 0$, then $S_n/\sigma\sqrt{n} \xrightarrow{\mathcal{L}} N(0,1)$.

Let $x_{1d}, ..., x_{Ld}, x'_{1d}, ..., x'_{Ld}$ be positive constants which, for the moment, are arbitrary. Let $A_{1d} = (0, x_{1d}) \times (0, x'_{1d}), \ldots, A_{L-1,d} = (0, x_{L-1,d}) \times (0, x'_{L-1,d})$ be different sets. Define

$$\bar{\Upsilon}_{A_{jd}} = \frac{1}{n}\sum_{i=1}^{n-1} I_{A_{jd}}(Y_{id}, Y_{i+1,d}), \tag{8.5}$$

where $I_A(.)$ is an indicator function. Then by the strong law of large numbers (SLLN), we have

$$\bar{\Upsilon}_{A_{jd}} \xrightarrow{a.s.} \Pr\{A_{jd}\} = \Pr\{Y_{id} \le x_{jd}, Y_{i+1,d} \le x'_{jd}\} \triangleq \mu_{jd}. \tag{8.6}$$

Let $\Upsilon_{nd} = I_{A_{jd}}\{Y_{nd}, Y_{n+1,d}\} - \mu_{jd}$, then $E[\Upsilon_{nd}] = 0$ and $E[\Upsilon_{nd}^{12}] < \infty$ because Υ_{nd} is bounded. The α-mixing condition is satisfied since Y_{nd}'s are M-dependent, i.e. Y_{id} and Y_{jd} are dependent when $|j-i| \le M$, while they are independent when $|j-i| > M$. So the conditions of **TA1** are satisfied. We have

$$\Upsilon_{1d}^2 = I_{A_{jd}}\{Y_{1d}, Y_{2d}\} - 2\mu_{jd} I_{A_{jd}}\{Y_{1d}, Y_{2d}\} + \mu_{jd}^2,$$

$$E\Upsilon_{1d}^2 = \mu_{jd} - 2\mu_{jd}^2 + \mu_{jd}^2 = \mu_{jd} - \mu_{jd}^2,$$

$$\begin{aligned}\Upsilon_{1d}\Upsilon_{1+k,d} &= (I_{A_{jd}}\{Y_{1d}, Y_{2d}\} - \mu_{jd})(I_{A_{jd}}\{Y_{1+k,d}, Y_{2+k,d}\} - \mu_{jd}) \\ &= I_{A_{jd}}\{Y_{1d}, Y_{2d}\} I_{A_{jd}}\{Y_{1+k,d}, Y_{2+k,d}\} \\ &\quad -\mu_{jd} I_{A_{jd}}\{Y_{1d}, Y_{2d}\} - \mu_{jd} I_{A_{jd}}\{Y_{1+k,d}, Y_{2+k,d}\} + \mu_{jd}^2\end{aligned}$$

and

$$E(\Upsilon_{1d}\Upsilon_{1+k,d}) = \Pr\{Y_{1d} \le x_{jd}, Y_{2d} \le x'_{jd}, Y_{1+k,d} \le x_{jd}, Y_{2+k,d} \le x'_{jd}\} - \mu_{jd}^2.$$

Then applying **TA1**, we have

$$\sqrt{n}(\bar{\Upsilon}_{A_{jd}} - \mu_{jd}) \xrightarrow{\mathcal{L}} N(0, \sigma_{jd}^2),$$

where μ_{jd} is the mean of random variable $\Upsilon_{A_{jd}}$. Its value is defined in (8.6). The value of σ_{jd}^2 is defined as:

$$\sigma_{jd}^2 = \mu_{jd} - \mu_{jd}^2 + 2\sum_{k=1}^{\max_l K_{1ld} + \max_l K_{2ld} + 1} \Big[\Pr\left\{Y_{1d} \le x_{jd}, Y_{2d} \le x'_{jd}, Y_{1+k,d} \le x_{jd}, Y_{2+k,d} \le x'_{jd}\right\} - \mu_{jd}^2\Big].$$

We now consider multivariate case. Let

$$U_{1d} = (I_{A_{1d}}\{Y_{1d}, Y_{2d}\} - \mu_{1d}, \ldots, I_{A_{L-1,d}}\{Y_{1d}, Y_{2d}\} - \mu_{L-1,d})',$$

$$U_{1+k,d} = (I_{A_{1d}}\{Y_{1+k,d}, Y_{2+k,d}\} - \mu_{1d}, \ldots, I_{A_{L-1,d}}\{Y_{1+k,d}, Y_{2+k,d}\} - \mu_{L-1,d})',$$

and $\alpha = (\alpha_{1d}, \ldots, \alpha_{L-1,d})' \neq 0$ be an arbitrary vector.

Let $\Upsilon_{1d} = \alpha' U_{1d}, \Upsilon_{2d} = \alpha' U_{2d}, \ldots$, then $E[\Upsilon_{nd}] = 0$ and $E[\Upsilon_{nd}^{12}] < \infty$. And so **TA1** can apply. We say expectation are applied on all elements if expectation is applied on a random matrix. But $E[\Upsilon_{1d}^2] = \alpha' E[U_{1d}U_{1d}']\alpha = \alpha'\Sigma\alpha$, $E[\Upsilon_{1d}\Upsilon_{1+k,d}] = \alpha' E[U_{1d}U_{1+k,d}']\alpha = \alpha' W_{kd}\alpha$ where

$$E([I_{A_{id}}\{Y_{1d}, Y_{2d}\} - \mu_{id}][I_{A_{jd}}\{Y_{1d}, Y_{2d}\} - \mu_{jd}]) = \mu_{i,j,d} - \mu_{id}\mu_{jd}$$

$$\begin{aligned} &E([I_{A_{id}}\{Y_{1d}, Y_{2d}\} - \mu_{id}][I_{A_{jd}}\{Y_{1+k,d}, Y_{2+k,d}\} - \mu_{jd}]) \\ &\quad = \Pr\{Y_{1d} \le x_{id}, Y_{2d} \le x_{id}', Y_{1+k,d} \le x_{jd}, Y_{2+k,d} \le x_{jd}'\} - \mu_{id}\mu_{jd} \end{aligned}$$

Applying the Cramér-Wold device, we have

$$\sqrt{n}\left(\begin{bmatrix} \bar{\Upsilon}_{A_{1d}} \\ \vdots \\ \bar{\Upsilon}_{A_{L-1,d}} \end{bmatrix} - \begin{bmatrix} \mu_{1d} \\ \vdots \\ \mu_{L-1,d} \end{bmatrix}\right) \xrightarrow{\mathcal{L}} N\Big(0, \Sigma_d + \sum_{k=1}^{\max_l K_{1ld} + \max_l K_{2ld}+1} \{W_{kd} + W_{kd}^T\}\Big)$$

where the entries $\sigma_{i,j,d}$ of matrix Σ_d are defined by: $\mu_{i,j,d} = \Pr\{Y_{1d} \le \min(x_{id}, x_{jd}),\ Y_{2d} \le \min(x_{id}', x_{jd}')\}$, $\sigma_{i,j,d} = \mu_{i,j,d} - \mu_{id}\mu_{jd}$, the matrix W_{kd} has entries $w_{kd}^{ij} = \Pr(Y_{1d} \le x_{id}, Y_{2d} \le x_{id}', Y_{1+k,d} \le x_{jd}, Y_{2+k,d} \le x_{jd}') - \mu_{id}\mu_{jd}$, $\mu_{i,i,d} = \mu_{id}$.

These arguments and the mean value theorem complete the proof of Lemma 3.1. □

Proof of Lemma 3.2. Since S is a set with finite number of distinct values, we have

$$C^* \in S^*,\ C_n^* \in S^*,\ n = 1, 2, \ldots$$

where $S^* = \{D_1, D_2, \ldots, D_T\}$ is a set of $l \times m$ matrices D_i whose elements belong to S, and T is finite. $D_i^T D_i$ is invertible.

Notice that $C_n^*\mathbf{a}_n^* \xrightarrow{a.s.} C^*\mathbf{a}^*$, as $n \to \infty$, implies that there is at least a subsequence $n_j,\ j = 1, 2, \ldots,$, such that $C_{n_j}^*\mathbf{a}_{n_j}^* \xrightarrow{a.s.} C^*\mathbf{a}^*$, as $j \to \infty$, and $C_{n_j} = D_i$ for some i, i.e. we have

$$C_{n_j}^*\mathbf{a}_{n_j}^* = D_i\mathbf{a}_{n_j}^* \xrightarrow{a.s.} C^*\mathbf{a}^*, \quad \text{as } j \to \infty$$

which implies that $D_i = C^*$, and $\mathbf{a}_{n_j}^* \xrightarrow{a.s.} \mathbf{a}^*$, as $j \to \infty$, and hence the proof is completed by noticing that for sufficiently large n, $C_n^* = C^*$. □

Acknowledgement

Zhengjun Zhang's work was supported in part by NSF grants DMS-0630210, DMS-0804575, Wisconsin Alumni Foundation, and SAMSI Risk Program 2007-2008. Richard Smith's work was supported in part by ..... The authors thank Rui Tang for pointing out an error in a proposition in one of our earlier versions of this paper.

References

[1] Beirlant, J., Geogebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley, Chichester.

[2] Berk, K. N. (1973) A central limit theorem for m-dependent random variables with unbounded m. *Annals of Probability*, **1**, 352-354.

[3] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, John & Wiley, New York.

[4] Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327

[5] Coles, S. G. (1993) Regional modelling of extreme storms via max-stable processes, *J. R. Statist. Soc. B,* **55**,pp. 797-816,

[6] Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer.

[7] Coles, S. G. and Tawn, J. A. (1991) Modeling extreme multivariate events. *J. R. Statist. Soc. B,* **53**,pp. 377-392

[8] Coles, S. G. and Tawn, J. A. (1994) Statistical methods for multivariate extrems: An application to structural design (with discussion). *Appl. Statist.* **43**, No. 1, pp. 1-48

[9] Coles, S. G. and Tawn, J. A. (1996) Modelling of extremes of the areal rainfall process. *J. R. Statist. Soc. B,* **58**,pp. 329-347

[10] Davis, R. A. and Resnick, S. I. (1989) Basic properties and prediction of Max-ARMA processes. *Adv. Appl. Prob.* , **21**, 781-803

[11] Davis, R. A. and Resnick, S. I. (1993) Prediction of stationary max-stable processes. *Ann. Appl. Prob.* , **3**, no 2, 497-525

[12] Davison, A.C. and Smith, R.L. (1990) Models for exceedances over high thresholds (with discussion). *J.R. Statist. Soc.* **52**, 393-442

[13] de Haan, L. (1984) A spectral representation for max-stable processes. *Ann. of Prob.* Vol 12, No. 4, 1194-1204

[14] de Haan, L. (1985) Extremes in higher dimensions: the model and some statistics. In *Proc. 45th Sess. Int. Statist. Inst.,* paper 26.3. The Hague: International Statistical Institute.

[15] de Haan, L. and Ferreira, A. *Extreme Value Theory: An Introduction.* Springer Verlag.

[16] de Haan, L. and Lin, T. (2003) Weak consistency of extreme value estimators in C[0,1]. *Annals of Statistics*, **31**, pp. 1996-2012.

[17] de Haan, L. and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete,* **40**, pp.317-337

[18] Deheuvels, P. (1978) Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publ. Inst. Statist. Univ. Paris. XXIII*, 1-36.

[19] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance,* Springer, Berlin.

[20] Embrechts, P., Lindskog, F. and McNeil, A. (2003) Modelling dependence with copulas and applications to risk management In: *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, Chapter 8, pp. 329-384

[21] Engle, R. (2002) Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*. Vol. 17, No. 5.

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

[23] Heffernan, J. and Tawn, J. (2004) A conditional approach for multivariate extreme values (with discussion). *J. Roy. Statist. Soc. B*, **66**, 497–546.

[24] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

[25] Ledford, A. W., Tawn, J. A., (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.

[26] Ledford, A. W., Tawn, J. A., (1997) Modelling dependence within joint tail regions, *J. Roy. Statist. Soc. B*, **59**, 475-499.

[27] Lee, S. -W., and Hansen, B. E. (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* 10, pp.29-52.

[28] Martins, A. P. and Ferreira, H. (2005) The multivariate extremal index and the dependence structure of a multivariate extreme value distribution. *Test*, **14**, 433-448.

[29] McNeil, A. and Frey, R. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: An extrem value approach. *Journal of Empirical Finance*, 7: 271-300

[30] Mikosch, T., (2003) Modeling depaendence and tails of financial time series. In *Extreme Value in Finance, Telecommunications and the Environment*, B. Finkenstadt and H. Rootzén, editors. Chapman and Hall/CRC.

[31] Pickands, J. III (1981) Multivariate extreme value distributions. *Proc. 43rd Session I.S.I.* (Buenos Aires), 859-878.

[32] Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

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

[34] Schlather, M. (2002) Models for stationary max-stable random fields. *Extremes* **5**, 33–44.

[35] Sibuya, M. (1960) Bivariate extreme statistics, I. *Ann. Inst. Statist. Math*, **11**, 195-210

[36] Smith, R. L. (1989) Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. (with discussion) *Statistical Science.* **4**, No 4, 367-393

[37] Smith, R.L. (2003) Statistics of extremes, with applications in the environment, insurance and finance. In *Extreme Value in Finance, Telecommunications and the Environment*, B. Finkenstadt and H. Rootzén, editors. Chapman and Hall/CRC.

[38] Smith, R. L., and Shively, T. S. (1995) Point process approach to modeling trends in tropo-sphereic ozone based on exceedances of a high threshold.*Atmospheric Environment,* **29**, No. 23, pp.3489-3499

[39] Smith, R. L., and Weissman, I. (1996) Characterization and estimation of the multivariate extremal index. Unpublished manuscript, University of North Carolina.

[40] Tsay, R. S. (1999) Extreme value analysis of financial data. Manuscript, University of Chicago.

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