Journal of Statistical Planning and Inference 140 (2010) 1135–1153

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On the estimation and application of max-stable processes

Zhengjun Zhang [a,*], Richard L. Smith [b]

[a] Department of Statistics, University of Wisconsin, Madison, WI 53706, USA
[b] Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, USA

ARTICLE INFO

Article history:
Received 29 January 2009
Received in revised form
27 October 2009
Accepted 28 October 2009
Available online 3 November 2009

Keywords:
Multivariate extremes
Multivariate maxima of moving maxima
Extreme value distribution
Empirical distribution
Estimation
Extreme dependence
Extreme co-movement

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.

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 et al. (1983), Embrechts et al. (1997), Coles (2001), Smith (2003), and most recently Gomes et al. (2008) 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),

* Corresponding author. Tel.: +1 608 262 2614; fax: +1 608 262 0032.
E-mail address: zjz@stat.wisc.edu (Z. Zhang).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.10.014

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 processes, 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 after marginal transformation 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 et al. (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 detailed technical proofs.

2. The model and its identifiability

Our starting point is a multivariate strongly stationary time series $\{X_{id}, i=0,\pm1,\pm2,\ldots, d=1,\ldots,D\}$, 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 $\Pr\{Y_{id}\le y\}=\exp(-1/y), 0<y<\infty$.

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

$$\Pr\{Y_{id}\le y_{id},\ i=i_1,\ldots,i_2,\ d=1,\ldots,D\}=[\Pr\{Y_{id}\le n y_{id},\ i=i_1,\ldots,i_2,\ d=1,\ldots,D\}]^n,\quad n\ge 1.$$

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.

In characterizing multivariate extremal indices and approximating max-stable 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 each observed component random variable being unit Fréchet may be approximated by a *multivariate maxima of moving maxima* process, or M4 for short, with the representation

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

where $\{Z_{l,i},\ l=1,2,\ldots,\ -\infty<i<\infty\}$ are independent unit Fréchet random variables and $a_{l,k,d}$ are nonnegative coefficients satisfying $\sum_{l=1}^{\infty}\sum_{k=-\infty}^{\infty}a_{l,k,d}=1$ for each d. We note that in the literature, Deheuvels (1983) introduced the moving minimum MM corresponding processes, and M4 process is a direct generalization to the multivariate case of Deheuvels (1983). Theorem 2 of Deheuvels (1983) establishes that for any finite n, the joint distribution of n consecutive values from a max-stable process with Fréchet marginals can be approximated by a weighted maximum $Y_{id}=\max_{-\infty<k<\infty}\alpha_{idk}Z_k$ for suitable α_{idk}, where Z_k's are IID unit Fréchet. Here "can be approximated" is in the sense that there is a sequence of such approximations that converges in distribution to the true distribution of the Y_{id}. We refer detailed arguments to Smith and Weissman (1996) as they are beyond the scope of this paper. In addition to the theories derived in Smith and Weissman (1996), Zhang (2009) develops new theories which suggest that M4 process models are a rich family which can be used in modeling a wide range of multivariate dependencies.

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 nonzero. Thus

we have the representation

$$Y_{id} = \max_{1\le l\le L} \max_{0\le k\le K} a_{l,k,d} Z_{l,i-k}, \quad -\infty < i < \infty, \ \ d = 1,\dots,D, \tag{2.1}$$

where L and K are finite and the coefficients satisfy $\sum_{l=1}^{L}\sum_{k=0}^{K} a_{l,k,d} = 1$ for each d. We note that in (2.1), some $a_{l,k,d}$'s may be zero. For example, $a_{l,k,d} = 0$ when $k < K_{1,l,d}$ or $k > K_{2,l,d}$, and $a_{l,k,d} > 0$ when $K_{1,l,d} \le k \le K_{2,l,d}$, for some parameters L_d, $K_{1,l,d}$ and $K_{2,l,d}$, $l = 1,\dots,L_d$, $d = 1,\dots,D$. Without loss of generality and for notation convenience, we simply use L and K in (2.1).

Probabilistic properties of the model (2.1) have been studied in Zhang and Smith (2004), Martins and Ferreira (2005), Heffernan et al. (2007), and Zhang (2009), among others. Heffernan et al. (2007) and Zhang (2009) demonstrate that M4 processes can model various extreme dependence structures. Zhang (2009) also demonstrates that the simplest M4 process performs as good as or better than the widely used Gumbel copula in modeling dependencies between bivariate random variables or among multivariate random variables, respectively.

We have a general joint probability formula (Zhang and Smith, 2004):

$$\Pr\{Y_{id} \le y_{id},\ 1 \le i \le r,\ 1 \le d \le D\} = \exp\left[-\sum_{l=1}^{L}\sum_{m=1-K}^{r} \max_{1-m\le k\le r-m} \max_{1\le d\le D} \frac{a_{l,k,d}}{y_{m+k,d}}\right], \tag{2.2}$$

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 $\Pr(Y_{id} \le y) = e^{-1/y}$, 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:

$$\Pr(Y_{id} \le y_{id}, Y_{i+1,d} \le y_{i+1,d}) = \exp\left[-\sum_{l=1}^{L}\sum_{m=1-K}^{2} \max\left\{\frac{a_{l,1-m,d}}{y_{id}}, \frac{a_{l,2-m,d}}{y_{i+1,d}}\right\}\right], \tag{2.3}$$

$$\Pr(Y_{id} \le y_{1d}, Y_{id'} \le y_{1d'}) = \exp\left[-\sum_{l=1}^{L}\sum_{m=1-K}^{1} \max\left\{\frac{a_{l,1-m,d}}{y_{1d}}, \frac{a_{l,1-m,d'}}{y_{1d'}}\right\}\right]. \tag{2.4}$$

Under model (2.1), an extremal event is typically a result of a single very large value of Z, say $Z_{l,i}$. When this happens, we will have $Y_{id} = a_{l,0,d} Z_{l,i}$, $Y_{i+1,d} = a_{l,1,d} Z_{l,i}, \dots, Y_{i+K,d} = a_{l,K,d} Z_{l,i}$. Thus in a local window, the sequence $\{Y_{id}\}$ as a vector is proportional to a deterministic sequence $a_{l,k,d}$ for some $l \in \{1,\dots,L\}$, which we call a signature pattern. Here L corresponds to the maximum number of distinct signature patterns. The constant K characterizes the range of dependence in each sequence and $(K+1)$ is the order of the moving maxima processes. We illustrate these phenomena in Fig. 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.

In the literature, Davis and Resnick (1985), Hsing (1986), and Rootzén (1986) investigated the limit cluster point processes from the extremes of moving average processes with regularly varying tails. They are excellent references in studying properties of extremes whose limits appear in clusters. The focuses in this paper are in statistical estimation of a parametric family of max-stable processes after marginal transformation to unit Fréchet scales.

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

Remark 1. For each l and d, the value of $\sum_{k=0}^{K} a_{l,k,d}$ 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 l th signature pattern.

For any pair of random variables Y_1 and Y_2 sharing the same marginal distribution with common upper endpoint $y_F \le \infty$, 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), a direct calculation of the tail dependence index $\lambda_{dd'}$ between Y_{id} and $Y_{id'}$ gives

$$\lambda_{dd'} = 2 - \sum_{l=1}^{L}\sum_{m=1-K}^{1} \max\{a_{l,1-m,d}, a_{l,1-m,d'}\}.$$

Moreover, for fixed d, $Y_{i_1 d}$ and $Y_{i_2 d}$ have positive tail dependence whenever $|i_1 - i_2| \le k (\le K)$, 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

$$\lambda_{d(k)} = 2 - \sum_{l=1}^{L}\sum_{m=1-K}^{1+k} \max\{a_{l,1-m,d}, a_{l,1+k-m,d}\}.$$

Heffernan et al. (2007) give more calculations for such processes. The tail dependence indices clearly show that M4 process models are suitable for random variables with extreme dependencies.

Notice that the joint distributions of random variables in an M4 process are singular, and density functions do not exist. For this reason, we do not apply the method of maximum likelihood, but instead propose estimators based on using (2.3) and (2.4) to approximate empirical bivariate distribution functions.

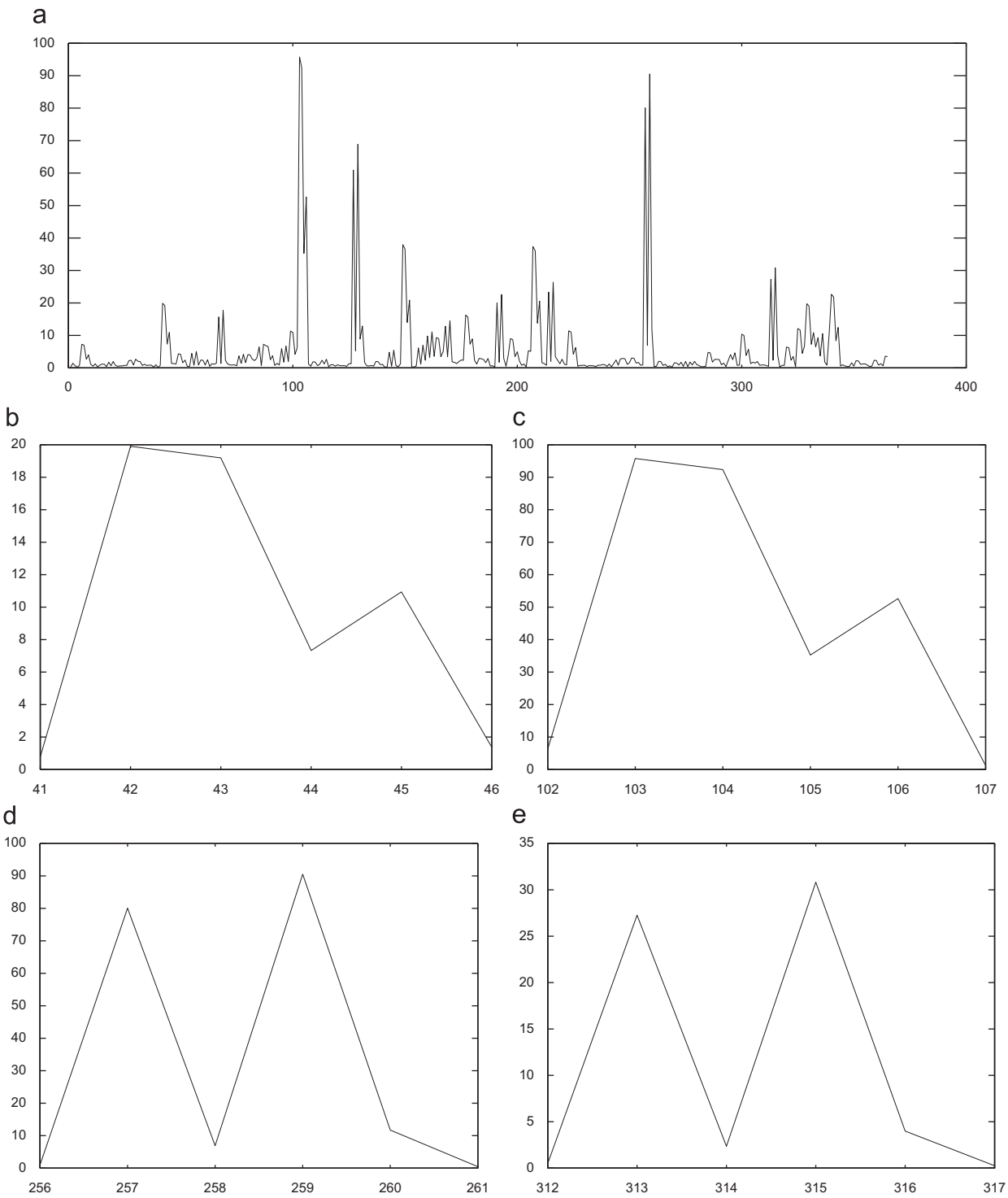

Fig. 1. A demonstration of an M4 process: (a) is a simulated 365 days data for a component process; (b)–(e) are partial pictures drawn from the whole simulated data showing two different moving patterns, called signature patterns, in certain time periods when extremal events occur.

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

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

So without loss of generality, we can fix $y_{id}=1$.

From (2.3) and (2.4), we have

$$\Pr\{Y_{id}\le 1, Y_{i+1,d}\le x\} = \exp\left[-\sum_{l=1}^{L}\sum_{m=1-K}^{2}\max\left(a_{l,1-m,d}, \frac{a_{l,2-m,d}}{x}\right)\right] =: e^{-b_{0d}(x)} \tag{2.6}$$

and

$$\Pr\{Y_{id}\le 1, Y_{id'}\le x\} = \exp\left[-\sum_{l=1}^{L}\sum_{m=1-K}^{1}\max\left(a_{l,1-m,d}, \frac{a_{l,1-m,d'}}{x}\right)\right] =: e^{-b_{dd'}(x)}. \tag{2.7}$$

It is clear that for each d, we can define new piecewise linear functions: $q_{0d}(x) := xb_{0d}(x)$ and $q_{dd'}(x) := xb_{dd'}(x)$, where the notation $A := B$ 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 $a_{l,k,d}/a_{l,k,d'}$. This suggests that if we can identify the functions $q_{0d}(x)$ or $q_{dd'}(x)$, we may be able to identify all the parameters $a_{l,k,d}$.

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 x values. It turns out we can do that and still retain identifiability, provided that the finite set of x values is chosen appropriately. The following important Proposition 2.1 will show the identifiability. It plays a core mathematical role in parameter estimation. Its proof is nonstandard and is deferred to Section 7.

Proposition 2.1. *Suppose the following three conditions hold*:

(i) *all nonzero and existing ratios $a_{l,j,d}/a_{l,j',d}$ for all l and $j\neq j'$ are distinct for each $d=1,\dots,D$;*
(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 $d=1,\dots,D$;*
(iii) *and nonzero and existing ratios $a_{l,k,1}/a_{l',k,d'}$ for all l, l' and k are distinct for each $d'=2,\dots,D$.*

Then $b_{0d}(x), d=1,\dots,D, b_{1d'}(x), d'=2,\dots,D$, uniquely determine all values of $a_{l,k,d}$, $d=1,\dots,D, l=1,\dots,L, 0\le k\le K$.

Furthermore, there exist points $x_1, x_2, \dots, x_m$, $m\le(2D-1)\sum_d\sum_l^L(K+2)+2D$, such that

$$b_{0d}(x_i) \quad \text{and} \quad b_{1d'}(x_i), \quad i=1,\dots,m, \ \ d=1,\dots,D, \ \ d'=2,\dots,D$$

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

Remark 2. When $D=1$ and $L=1$, the reason why Proposition 2.1 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 $q_{0d}(x)$. Condition (ii) assures that any permutation between different signature patterns will give a different $b_{0d}(.)$ 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 $a_{0,d},\dots,a_{K,d}$, the model is identifiable from $q_{0d}(x)$."

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 $0<k<K$, some of the ratios may be infinite, and not all coefficients a_{lkd} are identifiable by function $b_{0d}(x)$. This issue can be resolved by simply replacing $b_{0d}(x)$ by $b_{0d(r)}(x) = -\log[\Pr(Y_{id}\le 1, Y_{i+r,d}\le x)]$ for an appropriately chosen r, where $b_{0d(1)}(x)$ is just $b_{0d}(x)$ 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 2.1 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 $(a_0,\dots,a_4)=\frac{1}{6}(1,1,2,1,1)$ and $(b_0,\dots,b_4)=\frac{1}{6}(1,2,1,1,1)$. We consider the two processes generated by the sequences $a_0,\dots,a_4$ and $b_0,\dots,b_4$, i.e. $Y_i=\max_{k=0,1,2,3,4}a_kZ_{i-k}$,- $\infty<i<\infty$ and $Y_i'=\max_{k=0,1,2,3,4}b_kZ_{i-k}$, $-\infty<i<\infty$. For either process, the function $q(x)$ is piecewise linear with change points 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, they are distinguishable from trivariate distributions, and we have

$$-\log(\Pr(Y_1\le y_1, Y_2\le y_2, Y_3\le y_3)) = \frac{a_4}{y_1}+\max\left(\frac{a_3}{y_1},\frac{a_4}{y_2}\right)+\max\left(\frac{a_2}{y_1},\frac{a_3}{y_2},\frac{a_4}{y_3}\right)+\max\left(\frac{a_1}{y_1},\frac{a_2}{y_2},\frac{a_3}{y_3}\right)$$

$$+\max\left(\frac{a_0}{y_1},\frac{a_1}{y_2},\frac{a_2}{y_3}\right)+\max\left(\frac{a_0}{y_2},\frac{a_1}{y_3}\right)+\frac{a_0}{y_3}.$$

Taking $y_1=1, y_2=y_3=c>2$, we have $-\log(\Pr(Y_1\le y_1, Y_2\le y_2, Y_3\le y_3))=1+1/3c$, but $-\log(\Pr(Y_1'\le y_1, Y_2'\le y_2, Y_3'\le y_3))=1+1/2c$.

3. The estimators and asymptotics

3.1. Empirical distributions 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 $b_{0d}(x)$ and $b_{1d}(x)$ are defined as

$$U_{0d}(x)=\frac{1}{n-1}\sum_{i=1}^{n-1} I_{\{Y_{id}\le 1, Y_{i+1,d}\le x\}},\quad \hat{b}_{0d}(x)=-\log[U_{0d}(x)],\quad d=1,\dots,D, \tag{3.1}$$

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

where $I(\cdot)$ is an indicator function.

In the literature, the central limit theorem for m-dependent sequences is well established, for example, Berk (1973), 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. The following notations and their expressions will be used to establish Lemma 3.1. Let

$$x_{01d}, x_{02d}, \dots, x_{0m_0d},\quad d=1,\dots,D,$$

where $m_0\le(2D-1)(K+1)+2D$, and

$$x_{11d}, x_{12d}, \dots, x_{1m_1d},\quad d=2,\dots,D,$$

where $m_1\le(2D-1)\times D\times L\times(K+2)+2D$ 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:

$$\begin{aligned}
&\mathbf{x}=(x_{011},x_{021},\dots,x_{0m_01},x_{012},\dots,x_{0m_0D},x_{112},x_{122},\dots,x_{1m_12},x_{113},\dots,x_{1m_1D})^T,\\
&\mathbf{U}=(U_{01}(x_{011}),\dots,U_{01}(x_{0m_01}),U_{02}(x_{012}),\dots,\\
&\quad U_{0D}(x_{0m_0D}),U_{12}(x_{112}),\dots,U_{12}(x_{1m_12}),\dots,U_{1D}(x_{1m_1D}))^T,\\
&\hat{\mathbf{b}}=(\hat{b}_{01}(x_{011}),\dots,\hat{b}_{01}(x_{0m_01}),\hat{b}_{02}(x_{012}),\dots,\hat{b}_{0D}(x_{0m_0D}),\hat{b}_{12}(x_{112}),\\
&\quad \dots,\hat{b}_{12}(x_{1m_12}),\dots,\hat{b}_{1D}(x_{1m_1D}))^T.
\end{aligned}$$

We denote $\mathbf{b}$ as a vector of which the i th element is obtained by replacing the i th element of $\hat{\mathbf{b}}$ with its theoretical counterpart, and $\boldsymbol{\mu}$ as a vector of which the i th element is the expectation of the i th element of $\mathbf{U}$, denoted as

$$\mu_{ijd}:=E[U_{id}(x_{ijd})]=\Pr(Y_{1(d-i)^{1-i}}\le 1, Y_{2-i,d}\le x_{ijd}),\quad i=0,1,\ d=1+i,\dots,D,\ j=1,\dots,m_i.$$

Define

$$\begin{aligned}
w^{(k)}_{ijd,i'j'd'}&=E[(I_{\{Y_{1(d-i)^{1-i}}\le 1, Y_{2-i,d}\le x_{ijd}\}}-\mu_{ijd})(I_{\{Y_{1+k,(d'-i')^{1-i'}}\le 1, Y_{2+k-i',d'}\le x_{i'j'd'}\}}-\mu_{i'j'd'})]\\
&=\Pr(Y_{1(d-i)^{1-i}}\le 1, Y_{2-i,d}\le x_{ijd}, Y_{1+k,(d'-i')^{1-i'}}\le 1, Y_{2+k-i',d'}\le x_{i'j'd'})-\mu_{ijd}\mu_{i'j'd'},\quad i=0,1,\ d=i+1,\dots,D,\ j=1,\dots,m_i,\\
&\quad i'=0,1,\ d'=i'+1,\dots,D,\ j'=1,\dots,m_{i'}.
\end{aligned}$$

We use the following procedure to define $K+2$ matrices, i.e. $W_0, W_1, \dots, W_{K+1}$, and $w_k(r,s)$ is the r th row and the s th column element in the matrix W_k.

```
For k = 0 to K+1,
  For i = 0 to 1,
    For i' = 0 to 1,
      For d = i+1 to D,
        For d' = i'+1 to D,
          For j = 1 to m_i,
            For j' = 1 to m_i',
              r = (1 − i) × [(d − 1) × m_i + j] + i × [D × m_0 + (d − 2) × m_1 + j];
              s = (1 − i') × [(d' − 1) × m_i' + j'] + i' × [D × m_0 + (d' − 2) × m_1 + j'];
              w_k(r, s) = w^(k)_ijd,i'j'd';
```

and the matrix $\Theta=(\mathrm{diag}\{\boldsymbol{\mu}\})^{-1}\times(\mathrm{diag}\{\mathbf{x}\})$. Here both $\boldsymbol{\mu}$ and $\mathbf{x}$ are vectors, and the notation $\mathrm{diag}\{\boldsymbol{\mu}\}$ is to form a matrix with its diagonal elements being from $\boldsymbol{\mu}$. With the established notations, we now present one of our main results

in the following Lemma 3.1. The results in the lemma are for any arbitrary choices of x values in $\mathbf{x}$. 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.

Lemma 3.1. *For the choices of x values in* $\mathbf{x}$ *and with* $\boldsymbol{\mu}, \mathbf{b}, W_k, \Theta$ *defined above, we have*

$$\sqrt{n}(\mathbf{U}-\boldsymbol{\mu}) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, W_0 + \sum_{k=1}^{K+1}\{W_k + W_k^T\}\right)$$

and

$$\sqrt{n}(\hat{\mathbf{b}}-\mathbf{b}) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, \Theta\left[W_0 + \sum_{k=1}^{K+1}\{W_k + W_k^T\}\right]\Theta^T\right),$$

which establish the asymptotics for the empirical functions $\hat{b}_{0d}(x), \hat{b}_{1d'}(x),\ d=1,\ldots,D,\ d'=2,\ldots,D.$

Proof of Lemma 3.1 is deferred to Section 7.

3.2. Main results

The next step in the estimation procedure is that we use the estimated $\hat{\mathbf{b}}$ to solve for the $a_{l,k,d}$ values. Consider the system of nonlinear equations formed by (2.6) and (2.7), where x is now substituted by $x_{0jd},\ j=1,\ldots,m_0,\ d=1,\ldots,D$ in (2.6), and by $x_{1j'd'},\ j'=1,\ldots,m_1, d'=2,\ldots,D$ 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 Eqs. (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 $\{1, 1/x_{0jd}, 1+1/x_{0jd}, 1/x_{1j'd'},\ j=1,\ldots,m_0,\ d=1,\ldots,D,\ j'=1,\ldots,m_1,\ d'=2,\ldots, D\}$, and C^TC is invertible. Notice that (3.3) can also be expressed as $\mathbf{b} = CPP^T\mathbf{a}$, where P is a column permutation matrix and $PP^T = I$. We note that the existence of the permutation matrix P does not affect the estimation of the parameters in (2.1).

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

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

where each element in matrix $\hat{C}$ belongs to $\{1, 1/x_{0jd}, 1+1/x_{0jd}, 1/x_{1j'd'},\ j=1,\ldots,m,\ d=1,\ldots,D,\ j'=1,\ldots,m_1,\ d'=2,\ldots, D\}$, which is a finite set. The asymptotics of estimator $\hat{\mathbf{a}}$ is presented in the following main theorem whose proof is deferred to Section 7.

Theorem 3.2. *For the multivariate processes* $\{Y_{id}\}$, *suppose all three conditions in Proposition* 2.1 *are satisfied. Then there exist*

$$\{x_{01d}, x_{02d}, \ldots, x_{0m_0d}, d=1,\ldots,D\}$$

and

$$\{x_{11d'}, x_{12d'}, \ldots, x_{1m_1d'}, d'=2,\ldots,D\}$$

such that the estimator $\hat{\mathbf{a}} = (\hat{C}^T\hat{C})^{-1}\hat{C}^T\hat{\mathbf{b}}$ *satisfies*

$$\sqrt{n}(\hat{\mathbf{a}} - P^T\mathbf{a}) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, B\Theta\left[W_0 + \sum_{k=1}^{K+1}\{W_k + W_k^T\}\right]\Theta^T B^T\right),$$

where P *is column permutation matrix and* $B = P^T(C^TC)^{-1}C^T$.

Notice that the matrix $\hat{C}$ can be formed based on the vector $\mathbf{x}$ such that the first block of $\hat{\mathbf{a}}$ contains the estimates of $a_{l,k,1}$, and the second block of $\hat{\mathbf{a}}$ contains the estimates of $a_{l,k,2}$, and so on. Therefore, without loss of generality, we can assume P is a unit matrix.

We note that for each one-dimensional component process, as long as Conditions (i) and (ii) in Proposition 2.1 are satisfied, Theorem 3.2 is directly applicable, i.e. set $D=1$. However, when one combines all one-dimensional component processes for which Conditions (i) and (ii) in Proposition 2.1 are satisfied into an M4 process, there are some examples of M4 processes for which Condition (iii) is satisfied, but Conditions (i) and (ii) are not satisfied any more. A concrete example is in Section 5. For such cases, we propose an alternative methodology and use an example to illustrate the idea next.

3.3. A special treatment

Example 3.1. Suppose, for the d th 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 $\{a_{l,k,d},\ k=0,\dots,K\}$ are proportional to $\{a^*_{l^*,k,d},\ k=0,\dots,K\}$ 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

$$\begin{aligned} Y_{id} &= \max_{1\leq l\leq L}\max_{0\leq k\leq K} a_{l,k,d}Z_{l,i-k} \\ &= \max_{1\leq l\leq L}\beta_{ld}\max_{0\leq k\leq K} a^*_{l,k,d}Z_{l,i-k}, \quad d=1,\dots,D, \end{aligned} \tag{3.5}$$

where $(a^*_{l,k,d},\ k=0,\dots,K)=(a^*_{l',k,d},\ k=0,\dots,K)$ 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 $(a^*_{l,k,d},\ k=0,\dots,K)=(a^*_{l',k,d},\ k=0,\dots,K)$ which is related to a particular l^*.

This example says 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 $a_{l,j,d}/a_{l,j',d}$ are distinct. However, estimation of $a_{l,k,d}$ can be done via estimating $a^*_{l^*kd}$ and β_{ld} first, and the estimator is $\hat{a}_{l,k,d}=\hat{\beta}_{ld}\hat{a}^*_{l^*,k,d}$. Using the mean value theorem, the asymptotics of the estimator $\hat{\mathbf{a}}$ can be derived from the joint asymptotics of $\hat{\mathbf{a}}^*$ and $\hat{\beta}$.

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 3.3. *Suppose for each d in* (3.5), $a^*_{l,k,d}$*'s are estimated using observations from the dth component process. The limiting covariance matrices are denoted as* $\Sigma^*_1,\Sigma^*_2,\dots,\Sigma^*_D$*, respectively. Suppose the vector* $\mathbf{a}^*$ *consists of all the coefficients* $a^*_{l,k,d}$ *arranged in some order, and the corresponding estimators* $\hat{a}^*_{l,k,d}$ *are arranged in* $\hat{\mathbf{a}}^*$ *accordingly, then*

1. $\sqrt{n}(\hat{\mathbf{a}}^*-\mathbf{a}^*)\xrightarrow{\mathcal{L}}\mathrm{N}(0,\boldsymbol{\Sigma}^*)$, *where the diagonal matrices in* $\boldsymbol{\Sigma}^*$ *are* $\Sigma^*_1,\Sigma^*_2,\dots,\Sigma^*_D$.
2. *There is a matrix* $\boldsymbol{\Sigma}$ (*depending on* $\boldsymbol{\Sigma}^*$ *and the coefficients* $\{\beta_{l,d}\}$) *so that*

$$\sqrt{n}(\hat{\mathbf{a}}-\mathbf{a})\xrightarrow{\mathcal{L}}\mathrm{N}(0,\boldsymbol{\Sigma}),$$

where $\hat{a}_{l,k,d}=\beta_{ld}\hat{a}^*_{l^*,k,d}$, *and* $a_{l,k,d}=\beta_{ld}a^*_{l^*,k,d}$.

The proof of this corollary is simply due to the fact that $\hat{\mathbf{a}}$ coefficients are linear combinations of $\hat{\mathbf{a}}^*$ which has an asymptotic multivariate normal distribution. The arrangement does not depend on the sample size n. The form of $\boldsymbol{\Sigma}$ depends on particular identical signature patterns in $\{a_{l,k,d}\}$. The asymptotic standard deviations of $\{\hat{a}_{l,k,d}\}$ can be easily obtained from $\Sigma^*_1,\Sigma^*_2,\dots,\Sigma^*_D$, and $\beta_{l,d}$'s. In Section 5, we shall illustrate how to calculate asymptotic standard deviations using a real data example.

Theorem 3.2 and Corollary 3.3 establish the asymptotic distribution of the parameter estimators. In the next section we propose a procedure to determine the actual x_{ijd} used in the estimating equations.

4. Determining the x_{ijd} values and a simulation example

In real data, it is unlikely that we would observe exact replicates of a signature pattern as in Fig. 1 even after marginal transformations. There will either be measurement error (e.g. hydrology or meteorology data) or just random noise that cannot be captured by any theoretical model (e.g. the fact that prices on the stock exchange are continuously changing; they are obviously sensitive to exactly when the prices are recorded). As a result, any application of Model (2.1) to real data will first need marginally transforming data to unit Fréchet scales either by a parametric fitting method or a nonparametric method. Suppose the transformed variables are Y_{id}, then one may have

$$Y_{id}=Y^*_{id}+\varepsilon_{id},$$

where Y^*_{id} is a unit Fréchet random variable and ε_{id} is an error term which is normally distributed.

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. We note that the extremal properties of an M4 process are not changed by short-tailed additive noises, and we consider Y_{id} is approximately unit Fréchet distributed. 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 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 groups based on the consecutive ratios of $Y_{i+j-1,d}/Y_{i+j,d}$, $j=1,\dots,K$. The proposed procedure to determine the x_{0jd} and $x_{1j'd}$ values is then as follows:

1. For each d, use cluster analysis to group the consecutive ratios of $(Y_{i+j-1,d}/Y_{i+j,d},\ j=1,\dots,K)$ into L groups for all very large clustered observations indexed on i and appearing in K consecutive days. The tuning parameters L 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 $q_{0d}(x)$ changes slope. Between any two adjacent points, arbitrarily choose two points as x_{0jd} values. For example, suppose r_1, r_2 are two adjacent ratios, then a natural choice would be $x_{0jd}=r_1+0.25(r_2-r_1)$, $x_{0,j+1,d}=r_1+0.75(r_2-r_1)$.
4. The choices of $x_{1j'd}$ 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 $x_{1j'd}$ 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_{0jd} and $x_{1j'd}$ values, use them to estimate $a_{l,k,d}$ based on $\hat{b}_{0d}(x)$ and $\hat{b}_{1d}(x)$ functions.

Remark 4. In Step 1, we may need to cluster those very large clustered observations into more than L 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 $L+3$ groups. The three groups which have very small proportions among all those very large clustered observations are not used in determining x_{0jd} 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 Steps 3 and 4, theoretically, we should choose as many points of x_{0jd} and $x_{1j'd}$ 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

$$Y_{id}=\max_{1\le l\le 3}\max_{0\le k\le 2} a_{l,k,d}Z_{l,i-k}+N_{id},\quad -\infty<i<\infty,\ \ d=1,2, \tag{4.1}$$

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}, which can generally have a variance σ^2. 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_{0jd} and $x_{1j'd}$ values are determined using the procedure described earlier, and subsequently, values of $\hat{b}_{0d}(x_{0jd})$, and $\hat{b}_{1d}(x_{1j'd})$ can be estimated.

Now we have a system of nonlinear equations, whose variables are the values of $\hat{a}_{l,k,d}$. The estimates are found through a Monte Carlo optimization algorithm. For all l,k,d, we simulate 5000 vector values of $\hat{a}_{l,k,d}$'s from which the ratios $\hat{a}_{l,k+1,d}/\hat{a}_{l,k,d}$ and $\hat{a}_{l,k,1}/\hat{a}_{l,k,d}$ $(d>1)$ are falling in the regions determined by x_{0jd}, $x_{1j'd}$ computed in Steps 3 and 4. We keep the vector whose ratios $\hat{a}_{l,k+1,d}/\hat{a}_{l,k,d}$, $\hat{a}_{l,k,1}/\hat{a}_{l,k,d}$ 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 $b_{0d}(x)$ and $b_{1d}(x)$ computed using the kept $\hat{a}_{l,k,d}$ values to the estimated functions $\hat{b}_{0d}(x)$ and $\hat{b}_{1d}(x)$.

Table 1
Simulation results for model (4.1).

Parameter	True value	Estimated value	Standard deviation	Parameter	True value	Estimated value	Standard deviation
$a_{1,0,1}$	0.1500	0.0941	0.0418	$a_{1,0,2}$	0.0700	0.0379	0.0295
$a_{1,1,1}$	0.2000	0.1258	0.0351	$a_{1,1,2}$	0.0400	0.0231	0.0108
$a_{1,2,1}$	0.0200	0.0101	0.0486	$a_{1,2,2}$	0.0300	0.0166	0.0137
$a_{2,0,1}$	0.0500	0.0573	0.0573	$a_{2,0,2}$	0.1000	0.0949	0.0215
$a_{2,1,1}$	0.1000	0.1132	0.0558	$a_{2,1,2}$	0.1300	0.1387	0.0417
$a_{2,2,1}$	0.0300	0.0295	0.0456	$a_{2,2,2}$	0.1700	0.1765	0.0518
$a_{3,0,1}$	0.1600	0.2091	0.0459	$a_{3,0,2}$	0.1100	0.1272	0.0323
$a_{3,1,1}$	0.1700	0.2143	0.0610	$a_{3,1,2}$	0.1200	0.1325	0.0318
$a_{3,2,1}$	0.1200	0.1468	0.0582	$a_{3,2,2}$	0.2300	0.2527	0.0707

$x_{0j1}=(0.0775,\ 0.1550,\ 0.2582,\ 0.4120,\ 0.6163,\ 0.8058,\ 0.9806,\ 1.1363,\ 1.2728,\ 1.5032,\ 1.8274,\ 2.1884)$.
$x_{0j2}=(0.4421,\ 0.6287,\ 0.7071,\ 0.8318,\ 1.0027,\ 1.1423,\ 1.2505,\ 1.3050,\ 1.3057,\ 1.4641,\ 1.7802,\ 2.1321)$.
$x_{1j'd}=(0.1376,\ 0.2642,\ 0.4258,\ 0.5106,\ 0.5189,\ 0.5631,\ 0.6433,\ 0.7055,\ 0.7496,\ 0.9356,\ 1.2637,\ 1.4312,\ 1.4383,\ 1.6236,\ 1.9871,\ 2.8617,\ 4.2474,\ 5.4343)$.
Standard deviations are computed by applying Theorem 3.2.

In Table 1, the estimated values of almost all cases (except $a_{1,0,1}$, $a_{1,1,1}$) are very close to the true parameter values. The estimated values in cases of $a_{1,0,1}, a_{1,1,1}$ 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.

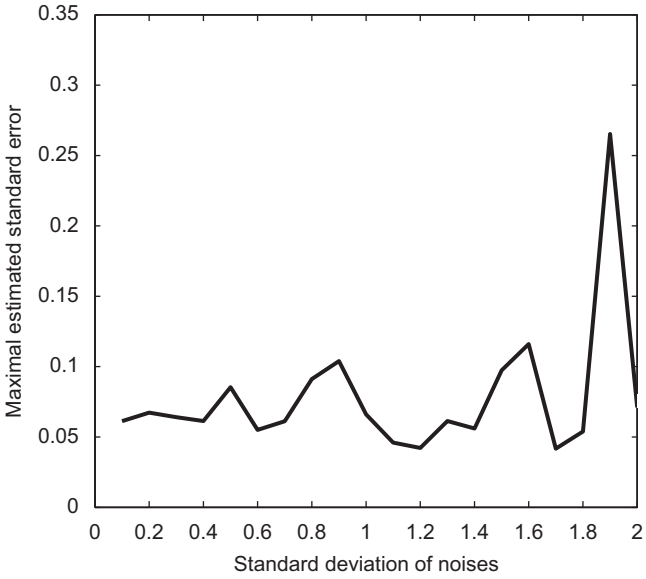

Fig. 2. Maximal simulated standard error of parameters versus standard deviations of N_{id}.

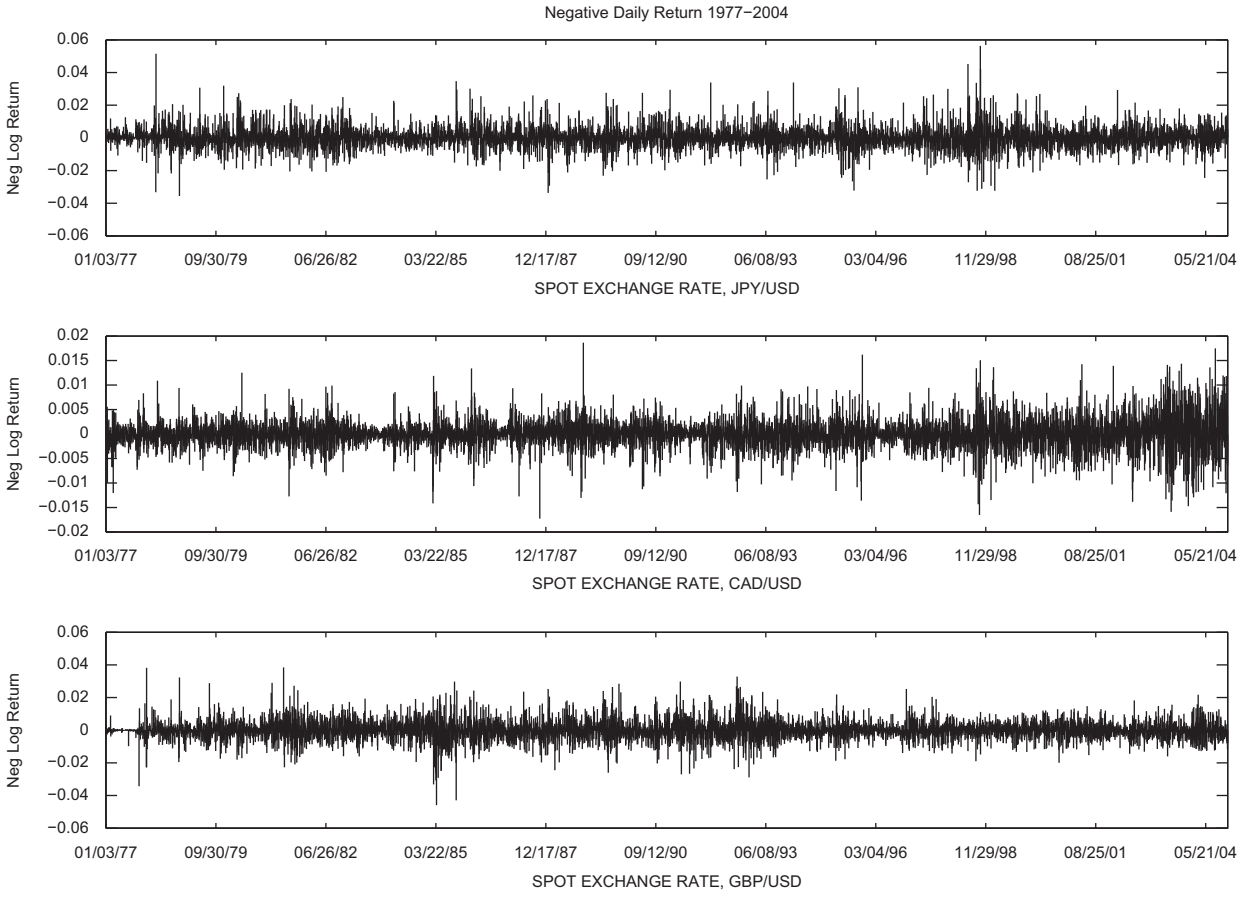

Fig. 3. Negative daily returns. The top plot is for JPY/USD, the middle plot is for CAD/USD, and the bottom plot is for GBP/USD.

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 $\sigma = 2$. For $\sigma > 2$, the optimization routines failed, mainly because of the initial clustering procedure. Fig. 2 depicts the maximum simulated standard error across all parameters as a function of σ. It illustrates that the standard error tends to increase with increasing σ.

5. Modeling jumps in returns of financial assets

We consider a trivariate time series of daily 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). Our data source is http://www.federalreserve.gov/Releases. The time range is from January 3, 1977 to December 31, 2004. They are plotted in Fig. 3.

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 Fig. 4. 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.

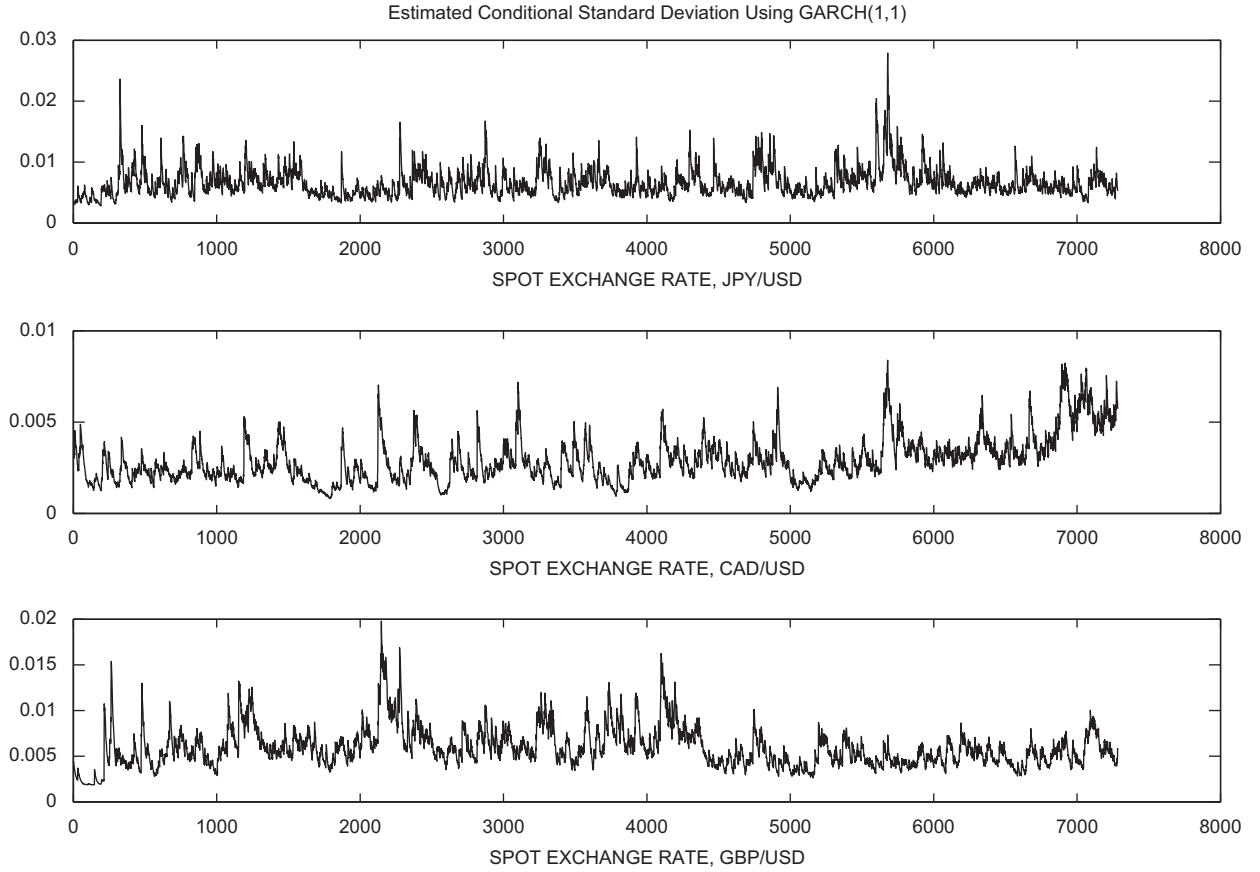

Fig. 4. Estimated volatility by GARCH. The top plot is for JPY/USD, the middle plot is for CAD/USD, and the bottom plot is for GBP/USD.

The 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 (0.02 for the original data, 1.2 for the volatility-standardized data) for each series. In each fitting, there are approximately 10% of observations which are above the chosen threshold value. 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 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.

We transform those observed values above the chosen threshold to unit Fréchet scales using $-1/\log(\hat{H}(x))$ where $\hat{H}(x)$ is the fitted GEV distribution function of (5.1) with the fitted GEV parameter values $(\hat{\mu}, \hat{\psi}, \hat{\xi})$, and then transform those observed values below the threshold using a rank based transformation method introduced in Coles and Tawn (1991), i.e. $-1/\log(\hat{F}(x))$. Here $\hat{F}(x)$ is an empirical distribution function. 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.

5.2. Model selection

Key parameters for determining the model are L, the number of clusters in the time series for component d, and the range parameter K. 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, Fig. 5 shows empirical tail dependence indices (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 indices suggest that random variables are tail dependent. 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.

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 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

Table 2
Estimation of parameters in GEV using standardized absolute return series.

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)

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.

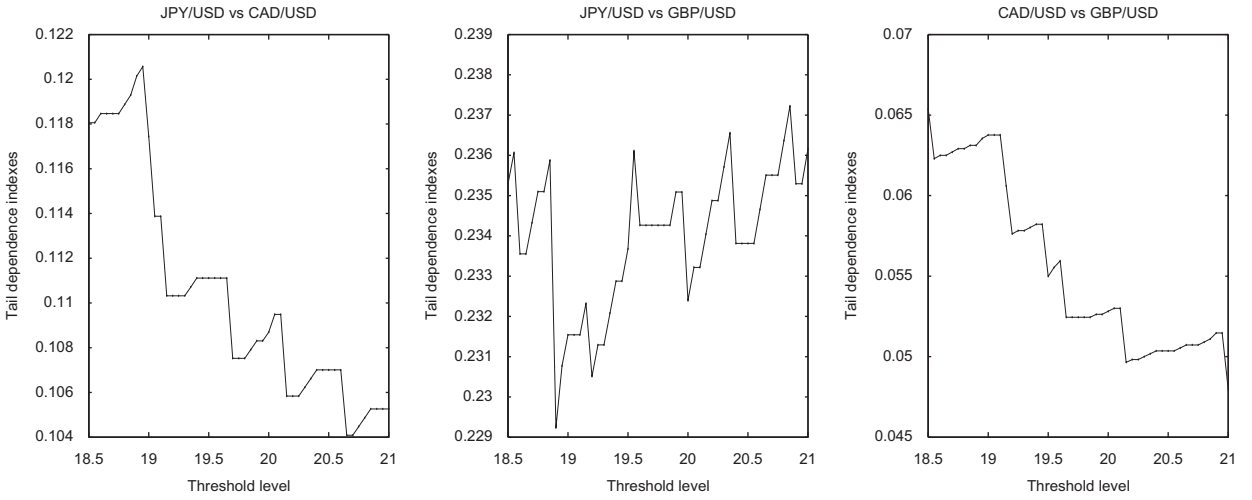

Fig. 5. Empirical estimates of tail dependence indexes against threshold values. The left panel (respectively, the middle panel, and the right panel) is for JPY/USD return and CAD/USD return (respectively, JPY/USD and GBP/USD, and CAD/USD and GBP/USD) at the same day.

Table 3
Counts of days that the absolute returns are over a threshold value in consecutive days.

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

All counts are mutually exclusive.

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 nonzero 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 3.2 is not directly applicable and we use Corollary 3.3 instead. If signature pattern 7 in Table 4 were not present, we would be able to use Theorem 3.2 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 3.2, therefore, we estimate parameters $a^*_{1,0,d}$, $a^*_{1,1,d}$, $a^*_{2,1,d}$, for each of $d = 1, 2, 3$. For example, in the case $d = 1$ we have $\hat{a}^*_{1,0,1} = 0.0165$, $\hat{a}^*_{1,1,1} = 0.0122$, $\hat{a}^*_{2,1,1} = 0.9713$. The full set of parameters $\{a_{l,k,d}\}$ is estimated as described in the discussion preceding Corollary 3.3. 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 three components and all 10 signature patterns in similar fashion, the full Table 4 is constructed.

After the model parameters are estimated, the next stage is to make statistical inferences based on the fitted model. Before one does the inferences, model fitting diagnoses are often needed in practice. For multivariate extreme value distribution fitting, one may use a nonparametric quantile–quantile (QQ) plot approach introduced in Genest and Rivest (1993), see also Zhang (2009). In this study, we propose to use an extreme co-movement measure to check the fitting next mainly because we are interested in market co-movements.

Table 4
Estimation of parameters in M4 model applied to standardized exchange rate time series.

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

The values in parentheses are standard errors.

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 (0.056)	0.9480 (0.037)	0.9535 (0.032)	0.9535 (0.032)	0.9535 (0.032)	0.9535 (0.032)	90th percentile
Model	0.9088 (0.008)	0.9446 (0.006)	0.9455 (0.006)	0.9455 (0.006)	0.9455 (0.006)	0.9455 (0.006)	
Data	0.8984 (0.088)	0.9805 (0.048)	0.9922 (0.031)	0.9922 (0.031)	0.9922 (0.031)	0.9922 (0.031)	97.5th percentile
Model	0.9425 (0.014)	0.9801 (0.008)	0.9811 (0.008)	0.9811 (0.008)	0.9811 (0.008)	0.9811 (0.008)	
Model	0.9629 (0.066)	0.9939 (0.025)	0.9956 (0.019)	0.9956 (0.019)	0.9956 (0.019)	0.9956 (0.019)	99.5th percentile

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 et al. 2003).

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. For example, the thresholds may be as high as 99.5th percentiles or even higher. 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 $\mathbf{u}$, and the inferences can go beyond the range of the data, 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.

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 2.1) 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 2.1 is easily extended to that case. Our main estimation results are in Theorem 3.2 and Corollary 3.3, 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 $x_{0jd}, x_{1j'd'}$ used in Theorem 3.2, 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. Technical arguments

Proof of Proposition 2.1. We first prove $q_{0d}(x)$ (or equivalently $b_{0d}(x)$ since $q_{0d}(x)=xb_{0d}(x)$) uniquely determine all coefficients in d th process. Since all the ratios are different, and they are points at which $q_{0d}(x)$ changes slopes, i.e. $q_{0d}'(x)$ has jumps, based on the jump points of $q_{0d}(x)$, the ratios of $a_{l,j+1,d}/a_{l,j,d}$ are uniquely determined. Let us now rewrite $q_{0d}(x)$ as

$$q_{0d}(x)=\sum_{l=1}^{L} xe_{ld} \sum_{m=1-K}^{2} \max\left(c_{l,1-m,d}, \frac{c_{l,2-m,d}}{x}\right), \tag{7.1}$$

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 $q_{0d}(x)$.

Suppose now $q_{0d}(x)$ has a different representation, say

$$q_{0d}(x)=\sum_{l=1}^{L} xf_{ld} \sum_{m=1-K}^{2} \max\left(c_{l,1-m,d}, \frac{c_{l,2-m,d}}{x}\right). \tag{7.2}$$

Then

$$\sum_{l=1}^{L}(e_{ld}-f_{ld}) \sum_{m=1-K}^{2} \max\left(c_{l,1-m,d}, \frac{c_{l,2-m,d}}{x}\right)=0 \tag{7.3}$$

for all x.

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

$$\Delta_d=\begin{bmatrix} \sum_{m=1-K}^{2} \max\left(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_1}\right) & \cdots & \sum_{m=1-K}^{2} \max\left(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_1}\right) \\ \vdots & \ddots & \vdots \\ \sum_{m=1-K}^{2} \max\left(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_{L-1}}\right) & \cdots & \sum_{m=1-K}^{2} \max\left(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_{L-1}}\right) \\ 1 & \cdots & 1 \end{bmatrix}$$

and set

$$\begin{bmatrix} \sum_{m=1-K}^{2} \max\left(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_1}\right) & \cdots & \sum_{m=1-K}^{2} \max\left(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_1}\right) \\ \vdots & \ddots & \vdots \\ \sum_{m=1-K}^{2} \max\left(c_{1,1-m,d}, \frac{c_{1,2-m,d}}{x_{L-1}}\right) & \cdots & \sum_{m=1-K}^{2} \max\left(c_{L,1-m,d}, \frac{c_{L,2-m,d}}{x_{L-1}}\right) \\ 1 & \cdots & 1 \end{bmatrix} \begin{pmatrix} e_{1d} - f_{1d} \\ \vdots \\ \vdots \\ e_{Ld} - f_{Ld} \end{pmatrix} = 0.$$

$|\Delta_d|$ is the determinant of the system of linear equations. Assume now the L determinants of the $(L-1)\times(L-1)$ matrices formed from the bottom $L-1$ rows are not all zero. Since $c_{l,k,d}$ are known and $\sum_{m=1-K}^{2} c_{l,i-m,d} = 1,\ i = 1,2$, 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 $1/x_1$ 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

$$|\Delta_d(x_1)| = \frac{1}{x_1}\sum c_{i,j,d}|\Delta_d|_{1j} + \sum c_{i',j',d}|\Delta_d|_{1j'}, \tag{7.4}$$

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 (7.4) are over all nonzero minors of the first row of Δ_d and the corresponding $c_{i,j,d}/x_1$ 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 $(1/x_1)c_{i,j,d}|\Delta_d|_{1j}$ of the summation $(1/x_1)\sum c_{i,j,d}|\Delta_d|_{1j}$ 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 $L\times L$ matrix such that the determinant is not zero as required. Therefore, there exist constants $x_1, \ldots, x_{L-1}$ such that each system of linear equations has a unique solution. We then conclude $e_{ld} = f_{ld}$, for all l. So $q_{0d}(x)$ uniquely determine all $a_{l,j,d}$.

Now, we prove the results for bivariate maxima and moving maxima processes. Since $b_{0d}(x)$ and $b_{0d'}(x)$ uniquely determine all values of parameters $a_{l,k,d}$ and $a_{l,k,d'}$, respectively, we can get

$$(a_{l,0,d}, a_{l,1,d}, \ldots, a_{l,K,d}), \quad l = 1, \ldots, L$$

and

$$(a_{l',0,d'}, a_{l',1,d'}, \ldots, a_{l',K,d'}), \quad l' = 1, \ldots, L.$$

Since all nonzero existing ratios $a_{l,k,d}/a_{l,k,d'}$ 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

$$\left(\frac{a_{l,0,d}}{a_{l,0,d'}}, \frac{a_{l,1,d}}{a_{l,1,d'}}, \ldots, \frac{a_{l,K,d}}{a_{l,K,d'}}\right)$$

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 $q_{0d}(x)$ 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 $D > 2$. □

The proof of Lemma 3.1 involves an $\alpha-$mixing condition and Theorem 27.4 in Billingsley (1995). They are re-stated below, and the latter is stated 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 $\alpha-$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 $\alpha-$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

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

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

Proof of Lemma 3.1. By the strong law of large numbers (SLLN), we have

$$U_{id}(x_{ijd}) \xrightarrow{a.s.} \mu_{ijd}, \quad i = 0, 1,\ \ d = i+1, \ldots, D,\ \ j = 1, \ldots, m_i. \tag{7.5}$$

Let $\Upsilon_{ind} = I_{\{Y_{n,(d-i)^{1-i}} \le 1, Y_{n+1-i,d} \le x_{ijd}\}} - \mu_{ijd}$ for any fixed j, then $E[\Upsilon_{nd}] = 0$ and $E[\Upsilon_{nd}^{12}] < \infty$ because Υ_{nd} is bounded. The $\alpha-$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

$$\Upsilon_{i1d}^2 = I_{\{Y_{1,(d-i)^{1-i}} \le 1, Y_{2-i,d} \le x_{ijd}\}} - 2\mu_{ijd} I_{\{Y_{1,(d-i)^{1-i}} \le 1, Y_{2-i,d} \le x_{ijd}\}} + \mu_{ijd}^2,$$
$$E\Upsilon_{i1d}^2 = \mu_{ijd} - 2\mu_{ijd}^2 + \mu_{ijd}^2 = \mu_{ijd} - \mu_{ijd}^2,$$

$$\begin{aligned}\Upsilon_{i1d}\Upsilon_{i,1+k,d} &= (I_{\{Y_{1,(d-i)^{1-i}} \le 1, Y_{2-i,d} \le x_{ijd}\}} - \mu_{ijd})(I_{\{Y_{1+k,(d-i)^{1-i}} \le 1, Y_{2+k-i,d} \le x_{ijd}\}} - \mu_{ijd}) \\ &= I_{\{Y_{1,(d-i)^{1-i}} \le 1, Y_{2-i,d} \le x_{ijd}\}} I_{\{Y_{1+k,(d-i)^{1-i}} \le 1, Y_{2+k-i,d} \le x_{ijd}\}} \\ &\quad - \mu_{ijd} I_{\{Y_{1,(d-i)^{1-i}} \le 1, Y_{2-i,d} \le x_{ijd}\}} - \mu_{ijd} I_{\{Y_{1+k,(d-i)^{1-i}} \le 1, Y_{2+k-i,d} \le x_{ijd}\}} + \mu_{ijd}^2\end{aligned}$$

and

$$E(\Upsilon_{i1d}\Upsilon_{i,1+k,d}) = w_{ijd,ijd}^{(k)}.$$

Then applying TA1, we have

$$\sqrt{n}(U_{id}(x_{ijd}) - \mu_{ijd}) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \sigma_{ijd}^2),$$

where σ_{ijd}^2 is defined as

$$\sigma_{ijd}^2 = \mu_{ijd} - \mu_{ijd}^2 + 2\sum_{k=1}^{K+1} w_{ijd,ijd}^{(k)}.$$

We now consider multivariate case. Let

$$U_{i,1+k,d}(x) = I_{\{Y_{1+k,(d-i)^{1-i}} \le 1, Y_{2+k-i,d} \le x\}}, \quad k = 0, 1, 2, \ldots$$

and column vectors whose construction is similar to $\mathbf{U}$ in Section 3.1:

$$\mathbf{U}_k^* = (U_{i,1+k,d}(x_{i,j,d}), \quad i = 0, 1, d = 1+i, \ldots, D, \quad j = 1, \ldots, m_i)' \quad k = 1, 2, \ldots.$$

Let α be an arbitrary nonzero vector whose length is the same as the length of $\mathbf{U}$.

Let $\Upsilon_1 = \alpha'(\mathbf{U}_1^* - \boldsymbol{\mu})$, $\Upsilon_2 = \alpha'(\mathbf{U}_2^* - \boldsymbol{\mu}), \ldots$, then $E[\Upsilon_n] = 0$ and $E[\Upsilon_n^{12}] < \infty$. And so TA1 can apply. We say expectation are applied on all elements if expectation is applied on a random matrix. But $E[\Upsilon_1\Upsilon_{1+k}] = \alpha' E[(\mathbf{U}_1^* - \boldsymbol{\mu})(\mathbf{U}_{1+k}^* - \boldsymbol{\mu})']\alpha = \alpha' W_k \alpha$ where the elements of W_k are

$$E([I_{\{Y_{1,(d-i)^{1-i}} \le 1, Y_{2-i,d} \le x_{ijd}\}} - \mu_{ijd}][I_{\{Y_{1+k,(d'-i')^{1+k-i'}} \le 1, Y_{2+k-i',d'} \le x_{i'j'd'}\}} - \mu_{i'j'd'}]) = w_{ijd,i'j'd'}^{(k)}.$$

Applying the Cramér–Wold device, we have

$$\sqrt{n}(\mathbf{U} - \boldsymbol{\mu}) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, W_0 + \sum_{k=1}^{K+1}\{W_k + W_k^T\}\right).$$

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

For notation convenience, we present the following definition:

Definition D1. For two vectors $\mathbf{b}$ and $\mathbf{a}$, and a matrix C, we say $\mathbf{b} = C\mathbf{a}$ up to a column permutation matrix P if $\mathbf{b} = CPP^T\mathbf{a}$. A sequence of random matrices C_n is said to converge to a matrix C almost surely up to a column permutation if $C_n \xrightarrow{a.s.} CP$, where P is a column permutation matrix. A sequence of column random vectors $\mathbf{a}_n$ is said to converge to a vector $\mathbf{a}$ almost surely up to a row permutation if $\mathbf{a}_n \xrightarrow{a.s.} P^T\mathbf{a}$, where P^T is a row permutation matrix. A q dimensional vector $\mathbf{a}$ with all elements being positive is said to be equivalent to a three-dimensional $(L \times (K+1) \times D)$ array $\tilde{\mathbf{a}}$ with all elements being nonnegative if all elements of $\tilde{\mathbf{a}}$ can be obtained in the following way:

```
(1)   t = 1;
(2)   For d = 1 to D;
        For l = 1 to L;
          For k = 1 to K+1;
            if ã_{l,k,d} > 0; ã_{l,k,d} = a(t); t=t+1;
```

The following Lemma 7.1 is important in proving our main Theorem 3.2.

Lemma 7.1. *Suppose S is a set with finite number of distinct values, $\mathbf{b}^*$ is a p dimensional vector, and suppose there is a unique $p \times q$, $(p \ge q)$, matrix C^* and a unique vector $\mathbf{a}^*$ such that $\mathbf{b}^* = C^*\mathbf{a}^*$ up to a column permutation, where the elements of C^* belong to S, $C^{*T}C^*$ being invertible; all elements $a^*(k)$ of $\mathbf{a}^*$ are positive and $\mathbf{a}^*$ is equivalent to a three-dimensional array $\tilde{\mathbf{a}}$ whose elements satisfy the three conditions of Proposition 2.1 and $\sum_{l,k}\tilde{a}_{l,k,d} = 1$, $d = 1, \ldots, D$. Suppose $\{\mathbf{b}_n^*, n = 1, 2, \ldots\}$ is a sequence of*

random vectors, $\{\mathbf{a}_n^*, n=1,2,\ldots\}$ *is also a sequence of random vectors with positive elements* $a_n^*(k)$ *and* $\mathbf{a}_n^*$ *is equivalent to a three-dimensional array* $\tilde{\mathbf{a}}_n$ *whose elements satisfy the three conditions of Proposition* 2.1 *and* $\sum_{l,k}\tilde{a}_{n,l,k,d}=1,\ d=1,\ldots,D$. *Moreover,* $\{C_n^*, n=1,2,\ldots\}$ *is a sequence of random matrices satisfying* $C_n^{*T}C_n^*$ *being invertible and each element of* C_n^* *belonging to* S. *Suppose* $\mathbf{b}_n^*=C_n^*\mathbf{a}_n^*,\ n=1,2,\ldots.\ \mathbf{b}_n^*\xrightarrow{a.s.}\mathbf{b}^*$, *as* $n\to\infty$, *then* $C_n^*\xrightarrow{a.s.}C^*P$, $\mathbf{a}_n^*\xrightarrow{a.s.}P^T\mathbf{a}^*$, *as* $n\to\infty$.

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

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

where $S^*=\{D_1,D_2,\ldots,D_T\}$ is a set of $p\times q$ matrices whose elements belong to S, and T is finite. $D_i^TD_i$ is invertible for each i.

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

Proof of Theorem 3.2. Since the estimators obey the strong law of large numbers, $\hat{\mathbf{b}}$ converges to $\boldsymbol{b}$ as $n\to\infty$, and there exists a unique column permutation matrix P such that

$$\begin{aligned}\sqrt{n}(\hat{\mathbf{a}}-P^T\mathbf{a}) &= \sqrt{n}((\hat{C}^T\hat{C})^{-1}\hat{C}^T\hat{\mathbf{b}}-P^T(C^TC)^{-1}C^T\mathbf{b})\\ &= \sqrt{n}(\hat{C}^T\hat{C})^{-1}\hat{C}^T(\hat{\mathbf{b}}-\mathbf{b})+\sqrt{n}((\hat{C}^T\hat{C})^{-1}\hat{C}^T-P^T(C^TC)^{-1}C^T)\mathbf{b}.\end{aligned}$$

By Lemma 7.1, $\hat{C}\xrightarrow{a.s.}CP$, $\hat{\mathbf{a}}\xrightarrow{a.s.}P^T\mathbf{a}$, as $n\to\infty$. The second term in the above equation disappears as n is sufficient large. Summarizing all arguments above, we have proved the theorem which is the asymptotic distribution of the estimators. □

Acknowledgments

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 NSF Grants DMS-0084375, DMS-0605434. The authors thank Rui Tang for pointing out an error in a proposition in one of our earlier versions of this paper. The authors also thank an associate editor and referees for their insightful comments and suggestions.

References

Beirlant, J., Geogebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. Wiley, Chichester.
Berk, K.N., 1973. A central limit theorem for m-dependent random variables with unbounded m. Ann. Probab. 1, 352–354.
Billingsley, P., 1995. Probability and Measure, third ed. Wiley, New York.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. J. Econometrics 31, 307–327.
Coles, S.G., 1993. Regional modelling of extreme storms via max-stable processes. J. Roy. Statist. Soc. B 55, 797–816.
Coles, S.G., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, Berlin.
Coles, S.G., Tawn, J.A., 1991. Modeling extreme multivariate events. J. Roy. Statist. Soc. B 53, 377–392.
Coles, S.G., Tawn, J.A., 1994. Statistical methods for multivariate extrems: an application to structural design (with discussion). Appl. Statist. 43 (1), 1–48.
Coles, S.G., Tawn, J.A., 1996. Modelling of extremes of the areal rainfall process. J. Roy. Statist. Soc. B 58, 329–347.
Davis, R.A., Resnick, S.I., 1985. Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. 13, 179–195.
Davis, R.A., Resnick, S.I., 1989. Basic properties and prediction of Max-ARMA processes. Adv. Appl. Probab. 21, 781–803.
Davis, R.A., Resnick, S.I., 1993. Prediction of stationary max-stable processes. Ann. Appl. Probab. 3 (2), 497–525.
Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds (with discussion). J. Roy. Statist. Soc. 52, 393–442.
de Haan, L., 1984. A spectral representation for max-stable processes. Ann. Probab. 12 (4), 1194–1204.
de Haan, L., 1985. Extremes in higher dimensions: the model and some statistics. In: Proceedings of the 45th Session of the International Statistical Institute, Paper 26.3. International Statistical Institute, The Hague.
de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer, Berlin.
de Haan, L., Lin, T., 2003. Weak consistency of extreme value estimators in C[0,1]. Ann. Statist. 31, 1996–2012.
de Haan, L., Resnick, S.I., 1977. Limit theory for multivariate sample extremes. Z. Wahrscheinlichkeitstheorie verw. Gebiete 40, 317–337.
Deheuvels, P., 1978. Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. Publ. Inst. Statist. Univ. Paris. XXIII, 1–36.
Deheuvels, P., 1983. Point processes and multivariate extreme values. J. Multivariate Anal. 13, 257–272.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.
Embrechts, P., Lindskog, F., McNeil, A., 2003. Modelling dependence with copulas and applications to risk management. In: Rachev, S. (Ed.), Handbook of Heavy Tailed Distributions in Finance. Elsevier, Amsterdam, pp. 329–384 (Chapter 8).
Engle, R., 2002. Dynamic conditional correlation—a simple class of multivariate GARCH models. J. Business Econom. Statist. 17 (5).
Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. J. Amer. Statist. Assoc. 88, 1034–1043.
Gomes, M.I., Canto e Castro, L.M., Isabel Fraga Alves, M.I.F., Pestana, D., 2008. Statistics of extremes for IID data and breakthroughs in the estimation of the extreme value index: Laurens de Haan leading contributions. Extremes 11, 3–34.
Hall, P., Peng, L., Yao, Q., 2002. Moving-maximum models for extrema of time series. J. Statist. Plann. Inference 103, 51–63.
Heffernan, J., Tawn, J., 2004. A conditional approach for multivariate extreme values (with discussion). J. Roy. Statist. Soc. B 66, 497–546.
Heffernan, J.E., Tawn, J.A., Zhang, Z., 2007. Asymptotically (in)dependent multivariate maxima of moving maxima processes. Extremes 10, 57–82.
Hsing, T., 1986. Extreme value theory for suprema of random variables with regularly varying tail probabilities. Stochastic Process. Appl. 22, 51–57.

Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer, New York.
Ledford, A.W., Tawn, J.A., 1996. Statistics for near independence in multivariate extreme values. Biometrika 83, 169–187.
Ledford, A.W., Tawn, J.A., 1997. Modelling dependence within joint tail regions. J. Roy. Statist. Soc. B 59, 475–499.
Lee, S.-W., Hansen, B.E., 1994. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory 10, 29–52.
Martins, A.P., Ferreira, H., 2005. The multivariate extremal index and the dependence structure of a multivariate extreme value distribution. Test 14, 433–448.
McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. J. Empirical Finance 7, 271–300.
Mikosch, T., 2003. Modeling dependence and tails of financial time series. In: Finkenstadt, B., Rootzén, H. (Eds.), Extreme Value in Finance, Telecommunications and the Environment. Chapman & Hall, CRC Press, London, Boca Raton, FL.
Pickands III, J., 1981. Multivariate extreme value distributions. In: Proceedings of the 43rd Session of the International Statistical Institute (Buenos Aires), pp. 859–878.
Resnick, S.I., 1987. Extreme Values, Regular Variation, and Point Processes. Springer, New York.
Resnick, S.I., 2007. Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer, New York.
Rootzén, H., 1986. Extreme value theory for moving average processes. Ann. Probab. 14, 612–652.
Schlather, M., 2002. Models for stationary max-stable random fields. Extremes 5, 33–44.
Sibuya, M., 1960. Bivariate extreme statistics, I. Ann. Inst. Statist. Math. 11, 195–210.
Smith, R.L., 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion). Statist. Sci. 4 (4), 367–393.
Smith, R.L., 2003. Statistics of extremes, with applications in the environment, insurance and finance. In: Finkenstadt, B., Rootzén, H. (Eds.), Extreme Value in Finance, Telecommunications and the Environment. Chapman & Hall, CRC Press, London, Boca Raton, FL.
Smith, R.L., Shively, T.S., 1995. Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold. Atmospheric Environ. 29 (23), 3489–3499.
Smith, R.L., Weissman, I., 1996. Characterization and estimation of the multivariate extremal index. Unpublished manuscript, University of North Carolina. 〈http://www.stat.unc.edu/postscript/rs/extremal.pdf〉.
Tsay, R.S., 1999. Extreme value analysis of financial data. Manuscript, University of Chicago.
Zhang, Z., 2009. On approximating max-stable processes and constructing extremal copula functions. Statist. Inference Stochastic Process. 12, 89–114.
Zhang, Z., Smith, R.L., 2004. The behavior of multivariate maxima of moving maxima processes. J. Appl. Probab. 41, 1113–1123.