A New Class of Tail-dependent Time Series Models and Its Applications in Financial Time Series

Zhengjun Zhang
Department of Mathematics,
Washington University,
Saint Louis, MO 63130-4899, USA

Abstract

In this paper, the gamma test is used to determine the order of lag-k tail dependence existing in financial time series. Using standardized return series, statistical evidence based on the test results show that jumps in returns are not transient. New time series models which combine a specific class of max-stable processes, Markov processes, and GARCH processes are proposed and used to model tail dependencies within asset returns. Estimators for parameters in the models are developed and proved to be consistent and asymptotically joint normal. These new models are tested on simulation examples and some real data, the S&P500.

Keywords: extreme value distribution, max-stable process, Markov chain, GARCH, nonlinear time series, tail dependence index, the gamma test, lag-k tail dependence, financial risk.

Journal of Economic Literature Classification System: C22, C12, C13.

1 Introduction

Tail dependence – which is also known as asymptotic dependence or extreme dependence – exists in many applications, especially in financial time series analysis. Not taking this dependence into account may lead to misleading results. Towards a solution to the problem, we show statistical evidences of tail dependencies existing in jumps in returns from standardized asset returns, and propose new nonlinear time series models which can be used to model serially tail dependent observations.

A natural way of modeling tail dependencies is to apply extreme value theory. It is known that the limiting distributions of univariate and multivariate extremes are max-stable, as shown by Leadbetter, Lindgren and Rootzén (1983) in the univariate case and Resnick (1987) in the multivariate case. Max-stable processes, introduced by de Haan (1984), are an infinite-dimensional generalization of extreme value theory, and they do have the potential to describe clustering behavior and tail dependence.

Parametric models for max-stable processes have been considered since the 1980s. Deheuvels (1983) defines the moving minima (MM) process. Davis and Resnick (1989) study what they call the max-autoregressive moving average (MARMA) process of a stationary process. For prediction, see also Davis and Resnick (1993). Recently, Hall, Peng, and Yao (2002) discuss moving maxima models. In the study of the characterization and estimation of the multivariate extremal index, introduced by Nandagopalan (1990, 1994), Smith and Weissman (1996) extend Deheuvels' definition to the so-called multivariate maxima of moving maxima (henceforth M4) process.

Like other existing nonlinear time series models, motivations of using max-stable process models are still not very clear. First, real data applications of these models are yet becoming available. Also, statistical estimation of parameters in these models is not easy due to the nonlinearity of the models and the degeneracy of the joint distributions. A further problem is that a single form of moving maxima models is not realistic since the clustered blowups of the same pattern will appear an infinite number of times as shown in Zhang and Smith (2004). In our paper, we explore the practical motivations of applying certain classes of max-stable processes, GARCH processes and Markov processes.

In time series modeling, a key step is to determine a workable finite dimensional representation of the proposed model, i.e. to determine statistically how many parameters have to be included in the model. In linear time series models, such as AR(p), MA(q), auto-covariance functions or partial auto-covariance functions are often used to determine this dimension, such as the values of p and q. But in a nonlinear time series model, these techniques may no longer be applicable. In the context of max-stable processes, since the underlying distribution has no finite variance, the dimension of the model can not be determined in the usual manner. We use the gamma test and the concept of lag-k tail dependence, introduced by Zhang (2003a), to accomplish the task of model selection. Comparing with popular model selection procedures, for example Akaike's (Akaike, 1974) Information Criteria (AIC), Schwarz Criteria (BIC) (1978) , our procedure first determines the dimension of the model prior to the fitting of the model.

In this paper, we introduce a base model which is a specific class of maxima of moving maxima processes (henceforth M3 processes) as our newly proposed nonlinear time series model. This base model is mainly motivated through the concept of tail dependencies within time series. Although it is a special case of the M4 process, the motivation and the idea are new. This new model not only possesses the properties of M4 processes, which have been studied in detail in Zhang and Smith (2004), but statistical estimation turns out to be relatively easy.

Starting from the basic M3 process at lag-k, we further improve the model to allow for possible asymmetry between positive and negative returns. This is achieved by combining an M3 process with an independent two state ({0,1}) Markov chain. The M3 process handles the jumps, whereas the Markov chain models the sign changes. In a further model, we combine two independent M3 processes, one for positive and negative within each, with an independent three state Markov chain ({-1, 0, 1}) governing the sign changes in the returns.

In Section 2, we introduce the concepts of lag-k tail dependence and the gamma test. Some theoretical results are listed in that section; proofs are given in Section 9. In Section 4, we introduce our proposed model and the motivations behind it. The identifiability of the model is discussed. The tail dependence indexes are explicitly derived under our new model. In Section 5, we combine the models introduced in Section 4 with a Markov process for the signs. In Section 7, we apply our approach to the S&P500 index. Concluding remarks are listed in Section 8. Finally, Section 9 continues the more technical proofs.

2 The concepts of lag-k tail dependence and the gamma test

Sibuya (1960) introduces tail independence between two random variables with identical marginal distributions. De Haan and Resnick (1977) extend it to the case of multivariate random variables. The definition of tail independence and tail dependence between two random variables is given below.

Definition 2.1 *A bivariate random variable* $(X_1,\ X_2)$ *is called tail independent if*

$$\lambda = \lim_{u \to x_F} P(X_1 > u | X_2 > u) = 0, \tag{2.1}$$

where X_1 *and* X_2 *are identically distributed with* $x_F = \sup\{x \in \mathbb{R} :\ P(X_1 \le x) < 1\}$*;* λ *is also called the bivariate tail dependence index which quantifies the amount of dependence of the bivariate upper tails. If* $\lambda > 0$*, then* $(X_1,\ X_2)$ *is called tail dependent.*

If the joint distribution between X_1 and X_2 is known, we may be able to derive the explicit formula for λ. For example, when X_1 and X_2 are normally distributed with correlation $\rho \in (0, 1)$ then, $\lambda = 0$. When X and Y have a standard bivariate t-distribution with ν degrees of freedom and correlation $\rho > -1$ then, $\lambda = 2\bar{t}_{\nu+1}(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$, where $\bar{t}_{\nu+1}$ is the tail of standard t distribution. Embrechts, McNeil, and Straumann (2002) give additional cases where the joint distributions are known.

Zhang (2003a) extends the definition of tail dependence between two random variables to lag-k tail dependence of a sequence of random variables with identical marginal distribution. The definition of lag-k tail dependence for a sequence of random variables is given below.

Definition 2.2 *A sequence of sample* $\{X_1,\ X_2,\ \ldots,\ X_n\}$ *is called lag-k tail dependent if*

$$\lambda_k = \lim_{u \to x_F} P(X_1 > u | X_{k+1} > u) > 0,\ \lim_{u \to x_F} P(X_1 > u | X_{k+j} > u) = 0,\ j > 1 \tag{2.2}$$

where $x_F = \sup\{x \in \mathbb{R} :\ P(X_1 \le x) < 1\}$*;* λ_k *is called lag-k tail dependence index.*

Here we need to answer two questions: the first is whether there is tail dependence; the second is how to characterize the tail dependence index. The first question can be put into the following testing of hypothesis problem between two random variables X_1 and X_2:

$$H_0: \; X_1 \text{ and } X_2 \text{ are tail independent} \leftrightarrow H_1: \; X_1 \text{ and } X_2 \text{ are tail dependent,} \tag{2.3}$$

which can also be written as

$$H_0: \; \lambda = 0 \; vs. \; H_1: \; \lambda > 0. \tag{2.4}$$

To characterize and test tail dependence is a difficult exercise. The main problem for computing the value of the tail dependence index λ is that the distribution is unknown in general; at least some parameters are unknown.

Our modeling approach first starts with the gamma test for tail (in)dependence of Zhang (2003a) which also demostrates that the gamma test efficiently detects tail (in)dependencies at high threshold levels for various examples. The test goes as follows.

Let

$$\begin{pmatrix} X_1, & X_2, & \dots, & X_n \\ Y_1, & Y_2, & \dots, & Y_n \end{pmatrix} \tag{2.5}$$

be an independent array of unit Fréchet random variables which have distribution function $F(x) = \exp(-1/x), \; x > 0.$. Now let $(U_i, \; Q_i), \; i = 1, \dots, n$ be a bivariate random sequence, where both U_i and Q_i are correlated and have support over $(0, u]$ for a typically high threshold value u. Let $X_{ui} = X_i I_{\{X_i > u\}} + U_i I_{\{X_i \le u\}}$, $Y_{ui} = Y_i I_{\{Y_i > u\}} + Q_i I_{\{Y_i \le u\}}$, $i = 1, \dots, n$. Then

$$\begin{pmatrix} X_{u1} \\ Y_{u1} \end{pmatrix}, \begin{pmatrix} X_{u2} \\ Y_{u2} \end{pmatrix}, \dots, \begin{pmatrix} X_{un} \\ Y_{un} \end{pmatrix} \tag{2.6}$$

is a bivariate random sequence drawn from two dependent random variables X_{ui} and Y_{ui}. Notice that $X_{ui} I_{\{X_{ui} > u\}}$ $(= X_i I_{\{X_i > u\}})$ and $Y_{ui} I_{\{Y_{ui} > u\}}$ $(= Y_i I_{\{Y_i > u\}})$ are independent; but $X_{ui} I_{\{X_{ui} \le u\}}$ $(= U_i I_{\{X_i \le u\}})$ and $Y_{ui} I_{\{Y_{ui} \le u\}}$ $(= Q_i I_{\{Y_i \le u\}})$ are dependent. Consequently, if only tail values are concerned, we can assume the tail values are drawn from (2.5) under the null hypothesis of tail independence. We have the following theorem.

Theorem 2.3 *Suppose V_i and W_i are exceedance values (above u) in (2.5). Then*

$$P\Big(\frac{u + W_i}{u + V_i} \le t\Big) = \begin{cases} \frac{t}{1+t} - \frac{t}{1+t} e^{-(1+t)/u}, & \text{if } 0 < t < 1, \\ \frac{t}{1+t} + \frac{1}{1+t} e^{-(1+t)/u}, & \text{if } t \ge 1, \end{cases} \tag{2.7}$$

$$\lim_{n \to \infty} P\Big(n^{-1}[\max_{i \le n} (u + W_i)/(u + V_i) + 1] \le x\Big) = e^{-(1 - e^{-1/u})/x}. \tag{2.8}$$

Moreover,

$$\lim_{n \to \infty} P\Big(n[\min_{i \le n} (u + W_i)/(u + V_i)] \le x\Big) = 1 - e^{-(1 - e^{-1/u})x}. \tag{2.9}$$

The random variables $\max_{i \le n} (u + W_i)/(u + V_i)$ *and* $\max_{i \le n} (u + V_i)/(u + W_i)$ *are tail independent, i.e.*

$$\lim_{n \to \infty} P\big(n^{-1}[\max_{i \le n} \frac{(u + W_i)}{(u + V_i)} + 1] \le x, \; n^{-1}[\max_{i \le n} \frac{(u + V_i)}{(u + W_i)} + 1] \le y\big) = e^{-(1 - e^{-1/u})/x - (1 - e^{-1/u})/y}. \tag{2.10}$$

Furthermore, the random variable

$$\mathrm{Q}_{u,n} = \frac{\max_{i \le n}\{(u+W_i)/(u+V_i)\} + \max_{i \le n}\{(u+V_i)/(u+W_i)\} - 2}{\max_{i \le n}\{(u+W_i)/(u+V_i)\} \times \max_{i \le n}\{(u+V_i)/(u+W_i)\} - 1}, \tag{2.11}$$

is asymptotically gamma distributed, i.e. for $x \ge 0$,

$$\lim_{n \to \infty} P(\mathrm{Q}_{u,n} \le x) = \zeta(x) \tag{2.12}$$

where ζ *is gamma*$(2, 1-e^{-1/u})$ *distributed.*

A proof of Theorem 2.3 is given in Zhang (2003a). Under certain mixing conditions, Zhang (2003a) also shows that (2.11) and (2.12) hold when V_i and W_i are not exceedances from (2.5).

Remark 1 *If* $V_i = (X_{ui} - u)I_{\{X_{ui}>u\}}$ *and* $W_i = (Y_{ui} - u)I_{\{Y_{ui}>u\}}$, *then (2.11) and (2.12) hold.*

Equations (2.11) and (2.12) together provide a gamma test statistic which can be used to determine whether tail dependence between two random variables is significant or not. When $n\mathrm{Q}_{u,n} > \zeta_\alpha$, where ζ_α is the upper αth percentile of the gamma$(2, 1-e^{-1/u})$ distribution, we reject the null-hypothesis of no tail dependence[1].

Remark 2 *In the case of testing for lag-k tail dependence, we let* $Y_i = X_{i+k}$ *in the gamma test statistic (2.11). If the observed process is lag-j tail dependent, the gamma test would reject the* H_0 *of (2.3) when* $k = j$, *but it would retain the* H_0 *of (2.3) when* $k = j+1$ *and a bivariate subsequence* $\{(X_{i_l},\ Y_{i_l}),\ l = 1, ...,\ i = 1, ...\}$*of data are used to compute the value of the test statistic, where* $i_l - i_{l-1} > j$. *In Section 4, we will introduce a procedure using local windows to accomplish this testing procedure.*

So far, we have established testing procedure to determine whether there exists tail dependence between random variables. Once the H_0 of (2.3) is rejected, we would like to model the data using a tail dependence model. In the next section we explore how the existence of tail dependence may imply transient behaviors in the jumps of financial time series.

3 Jumps in returns are not transient: evidences

The S&P500 index has been analyzed over and over again, but models using extreme value theory to analyze S&P500 index returns are still somehow delicate. There are several references; examples include Tsay (1999), Daníelsson (2002). Models addressing tail dependencies within the observed processes are more difficult to find in the literature. In this section, we use the gamma test to check whether there exists tail dependence for the S&P500 return data.

Recall that, in order to use the gamma test for tail independence, the original distribution function needs to be transformed into unit Fréchet distribution. This transformation preserves the tail dependence parameter.

[1] Matlab codes for the gamma test are available upon request.

3.1 Data transformation and standardization

In Figure 1, we plot the negative log returns of the S&P500 index. The data are from July 3, 1962 to December 31, 2002. The highest value corresponds to Oct. 19, 1987 Wall Street crash. There clearly are large moves (possibly jumps) both in the returns as well as in the volatilities. We are especially interested in the tail dependence resulting from these large moves.

Figure 1 about here

As GARCH models have been quite successful in modelling volatilities in financial time series, we apply a GARCH fitting to the data. A GARCH(1,1) (Bollerslev, 1986) model fitted to the data is plotted in Figure 1. The estimated volatility process is plotted in Figure 2. The original data, divided by the estimated standard volatility, are plotted in Figure 3. This series is referred as the GARCH-devolatilized time series. Examples of papers going beyond this approach are for instance McNeil and Frey (2000), and Engle (2002). In this paper, we study the standardized financial time series, or GARCH residuals using our methods. We focus on the modeling of the larger jumps in the standardized returns.

Figure 2 about here

Figure 3 about here

Notice that the largest value in the standardized series does not correspond to Oct. 19, 1987 Wall Street crash. The larger values in the standardized series look more like extreme observations, not outliers. An example of simulated extreme process which demonstrates extreme observations is illustrated in a later section. We now turn to study data transformation methods.

Methods for exceedance modeling over high thresholds are widely used in the applications of extreme value theory. The theory used goes back to Pickands (1975), below we give a brief review. For full details, see Embrechts, Klüppelberg, and Mikosch (1997) and the references therein. Consider the distribution function of a random variable X conditionally on exceeding some high threshold u, we have

$$F_u(y) = P(X \leq u + y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)},\ y \geq 0.$$

As $u \to x_F = \sup\{x :\ F(x) < 1\}$, Pickands (1975) shows that the generalized Pareto distributions (GPD) are the only non-degenerate distribution that approximate $F_u(y)$ for u large. The limit distribution of $F_u(y)$ is given by

$$G(y; \sigma, \xi) = 1 - (1 + \xi \frac{y}{\sigma})_+^{-1/\xi}. \tag{3.1}$$

In the GPD, $y_+ = \max(y, 0)$, ξ is the tail index, also called shape parameter, which gives a precise characterization of the shape of the tail of the GPD. For the case of $\xi > 0$, it is long-tailed, i.e., $1 - G(y)$ decays at rate $y^{-1/\xi}$ for large y. The case $\xi < 0$ corresponds to a distribution function which has a finite upper end point at $-\sigma/\xi$. The case $\xi = 0$ yields the exponential distribution with mean σ:

$$G(y; \sigma, 0) = 1 - \exp(-\frac{y}{\sigma}).$$

The Pareto, or GPD, and other similar distributions have long been used models for long-tailed processes. As explained in the previous sections, we first transform the data to unit Fréchet margins. This is done using the generalized extreme value (GEV) distributions which are closely related to the GPD as explained in Embrechts *et. al.* (1997) and can be written as:

$$H(x) = \exp\big[- (1 + \xi \frac{x-\mu}{\psi})_+^{-1/\xi} \big], \tag{3.2}$$

where μ is a location parameter, $\psi > 0$ is a scale parameter, and ξ is a shape parameter similar to the GPD form in (3.1). Pickands (1975) first established the rigorous connection of the GPD with the GEV. Using the GEV and GPD, we fit the tails of the negative observations, the positive observations, and the absolute observations separately. The absolute returns are included in our data analysis for comparison. The generalized extreme value model (3.2) is fitted to observations which are above a threshold u. We have tried a series of threshold values and performed graphical diagnosis using the mean excess plot, the W and Z statistics plots due to Smith (2003). Those diagnosis suggested that when values above $u = 1.2$, visually, a GPD function fits data well. There are about 10% of observed values above the threshold $u = 1.2$. The maximum likelihood estimates are summarized in Table 1. From the table, we see that standardized negative returns are still fat tailed since the estimated shape parameter value is positive, but the standardized positive returns are now short tailed. The absolute returns are still fat tailed.

Table 1 about here

The standardized data are now converted into unit Fréchet scale for the observations above $u = 1.2$ using the tail fits resulting from Table 1. The final transformed data are plotted in Figure 4. After standardization and scale transformation, we can say that the data show much less jumps in volatility, but jumps in returns are still persistent as will be shown in the next section.

Figure 4 about here

3.2 Tail dependence

In the applications of GARCH modelling, GARCH residuals are assumed to be independently identically distributed normal (or t) random variables. The residual distribution assumptions may not be appropriate in some applications. As a result, there have been various revised GARCH type models proposed in the literature. Hall and Yao (2003), Mikosch and Stărică (2000), Mikosch and Straumann (2002) – among others – have discussed heavy-tailed errors and extreme values. Straumann (2003) deals with the estimation in certain conditionally heteroscedastic time series models, such as GARCH(1,1), AGARCH(1,1), EGARCH etc. Our main interest is the analysis of tail dependence within the standardized unit Fréchet transformed returns. Given such tail dependence, we want to present a relevant time series model capturing that behavior.

In order to check whether there exists tail dependence within the sequences in Figure 4, we perform the gamma test from Section 2 to the transformed returns with the threshold level at the 95th percentile (90th percentile is related to $u = 1.2$) of observed sequence as in Zhang (2003a).

However, in reality or even in simulated examples as shown in Section 4, one single test may not be enough to discover the true tail dependence or independence. As pointed out in Zhang (2003a), the gamma test is able to detect tail dependence between two random variables even if there are outliers in one of two observed sequences. In our case, we have a univariate sequence. If there are outliers in the sequence, a Type I error may occur since outliers cause smaller $\mathrm{Q}_{u,n}$ values in (2.11). If there are lower lag tail dependencies, a Type II error may occur since there are dependencies in each projected sequence. In order to minimize the probability of making errors, we select observations in local windows of size 500 to perform the gamma test. The observations in each window are used to compute the value of the gamma test statistic and a global conclusion is suggested at level $\alpha = .05$. Besides the gamma test, we also compute the empirical estimation of the tail dependence indexes. The empirical estimation is done by using the following procedure.

Empirical estimation procedures: Suppose $x_1, x_2, \ldots, x_n$ is a sequence of observed values and x^* is the 95th sample percentile. Then the empirical estimate of lag-k tail dependence index is $\sum_{i=1}^{n-k} I_{(x_i > x^*, x_{i+k} > x^*)} / \sum_{i=1}^{n-k} I_{(x_i > x^*)}$, where $I_{(.)}$ is an indictor function. The test results for lags 1–15 are summarized in Table 2.

Table 2 about here

Columns 2, 6 ($\times 100$) yield percentages of rejection of H_0 from all fully enumerated local windows of size 500. Of each data point, at least one component is nonzero. Columns 3, 7 are empirically estimated lag-k tail dependence indexes over a threshold value computed at the 95th percentile for the whole data. Columns 4, 8 are the minima of all computed $\mathrm{Q}_{u,n}$ values using (2.11) in all local windows. Columns 5, 9 are the maxima of all computed $\mathrm{Q}_{u,n}$ values using (2.11). The number .7015 in Column 2 means that if we partition the whole data set into 100 subsets, and perform the gamma test to each subset, there are about 70 subsets from which we would reject the tail independence hypothesis. The number .0700 in Column 3 is the empirically estimated tail dependence index which tells that when a large price drop is observed, and the resulting negative return is below the 95th percentile of the historical data, there is 7% chance to observe a large price drop in the kth day. The rest of numbers in the table can be interpreted similarly.

From Column 2 in Table 2, we see that for lags from 1 to 12, the percentages of rejecting H_0 are high for negative returns. At lags 1, 7, 10, 11, 12, tail independence is always rejected in the case of negative returns. In Column 3, if adding up all estimated lag-1 to lag-12 tail dependence indexes, the total percentage is over 30% which tells that there is more than 30% of chance that a large price drop may happen in one of the following 12 days. This should be seriously taken into account in modeling. Lag-13 is a break point at which the rejection rate of tail independence is 0 for negative returns. The empirically estimated dependence indexes in Column 3 show a decreasing trend at the beginning five lagged days; then after lag-5, they become small and no trend. The table also shows that each of the first five empirically estimated tail dependence indexes is falling in the range of the minima of $\mathrm{Q}_{u,n}$ value, and the maxima of $\mathrm{Q}_{u,n}$. These numbers suggest that there is tail dependence within the negative returns. All the empirically computed values, the $\mathrm{Q}_{u,n}$ values can be used as estimates of the tail dependence index. In Section 4, we compute theoretical lag-k tail dependence indexes for certain models.

Columns 5 – 8 are for positive returns, and can be interpreted in a similar way to negative returns. Column 5 shows that up to lag-8, the positive returns are tail dependent. Lag-5 tail independence is always rejected. Lag-9 is a break point at which the rejection rate of tail independence is 0. Like negative returns, Column 6 shows a decreasing trend of the empirically estimated tail dependence indexes. After lag-5, there is no pattern. Columns 7,8 are the minima and the maxima of empirical tail dependence index values from formula (2.11). Each of the first five numbers in Column 6 is falling in the range of the minima and the maxima in Columns 7, 8 with the same lag numbers.

In the literature, that jumps in returns are transient often refers to that a large jump (positive or negative)) does not have a future impact. When jumps in return series are transient, they are tail independent. Models dealing with transience of jumps in return series are for instance Eraker, Johannes, and Polson (2003), Duan, Ritchken, and Sun (2003). Our results (jumps in returns being tail dependent) suggest that jumps in returns are not transient. When jumps in returns are tail dependent, a large jump may have an extreme impact on the future returns, and should be taken into account in modeling.

In order to find a criterion for deciding on the maximal lag of which tail dependence exists, we propose the following procedure:

Empirical criterion: When the rejection rate of the lag-r test first reaches 0 or a rate substantially smaller than the rejection rate of the lag-$(r-1)$ test, and the estimated lag-r tail dependence index is less than 1%, let $k = r$.

At the moment, this test procedure is still purely *ad hoc*; more work on its statistical properties is needed in the future. Based on this criteria, however, we use lag-12 tail dependence for the transformed negative returns and lag-7 tail dependence for the transformed positive returns in the sections below.

The above test may suggest the existence of tail dependence; finding on appropriate time series model explaining these dependencies is another task. In the next section we introduce a versatile class of models aimed at achieving just that.

4 The basic model and the computation of lag-k tail dependence indexes

Suppose the maximal lag of tail dependence within a univariate sequence is K, and $\{Z_{li},\ l = 1,\dots,L,\ -\infty < i < \infty\}$ is an independent array, where the random variables Z_{li} are identically distributed with a unit Fréchet distribution function. Consider the model:

$$Y_i = \max_{0\le l\le L}\ \max_{0\le k\le K} a_{lk} Z_{l,i-k}, \quad -\infty < i < \infty, \tag{4.1}$$

where the constants $\{a_{lk}\}$ are nonnegative and satisfy $\sum_{l=0}^{L}\sum_{k=0}^{K} a_{lk} = 1$. When $L = 0$, model (4.1) is a moving maxima process. It is clear that the condition $\sum_{l=0}^{L}\sum_{k=0}^{K} a_{lk} = 1$ makes the random variable Y_i a unit Fréchet random variable. So Y_i can be thought of as being represented by an independent array of random variables which are also unit Fréchet distributed. The main idea behind (4.1) is that stochastic processes (Y_i) with unit Fréchet margins can be thought of as being represented (through (4.1)) by an infinite array of independent unit Fréchet random variables. A next step concerns the quality of such an approximation for finite (possibly small) values of L and K. The idea and the approximation theory are presented in Smith and Weissman (1996), Smith(2003).

Zhang (2004) establishes conditions for using a finite moving range M4 process to approximate an infinite moving range M4 process.

Under model (4.1), when an extreme event occurs or when a large Z_{li} occurs, $Y_i \propto a_{l,i-k}$ for $i \approx k$, i.e. if some Z_{lk} is much larger than all neighboring Z values, we will have $Y_i = a_{l,i-k} Z_{lk}$ for i near k. This indicates a moving pattern of the time series, known as signature pattern. Hence L corresponds to the maximal number of signature patterns. The constant K characterizes the range of dependence in each sequence and the order of moving maxima processes. We illustrate these phenomena for the case of $L = 0$ in Figure 5. Plots (b) and (c) involve the same values of $(a_{00},\ a_{01},\ \dots,\ a_{0K})$. Plot (b) is a blowup of a few observations of the process in (a) and Plot (c) is a similar blowup of a few other observations of the process in (a). The vertical coordinate scales of Y in Plot (b) are from 0 to 5, while the vertical coordinate scales of Y in Plot (c) are from 0 to 50. These plots show that there are characteristic shapes around the local maxima that replicate themselves. Those blowups, or replicates, are known as signature patterns.

Figure 5 about here

Remark 3 *: Model (4.1) is a simplified model of a general M4 process since we restrict our attention to univariate time series. In Zhang and Smith (2003, 2004), properties of models with a finite number of parameters are discussed. They provide sufficient conditions from which the estimators of model parameters are shown to be consistent and jointly asymptotic multivariate normal.*

Under model (4.1), we have the following lag-k tail dependence index formula:

$$\lambda_k = \sum_{l=1}^{\infty} \sum_{m=-\infty}^{\infty} \min(a_{l,1-m},\ a_{l,1+k-m}). \tag{4.2}$$

Obviously, as long as both a_{l0} and a_{lK} are non-zero, Y_i and Y_{i+K} are dependent, and of course tail dependent as can be seen from (4.2).

In real data, for example, the negative returns, we count how many times that negative return values $Y_i, Y_{i+1}, \dots, Y_{i+k}, (i = 1, \dots, n-k)$, are simultaneously greater than a given threshold value for any fixed k. We record all the i values such that $Y_i, Y_{i+1}, \dots, Y_{i+k}$ are simultaneously greater than the given threshold value . We typically find that jumps in negative returns as well as jumps in positive returns appear to cluster in two consecutive days, i.e. $k = 1$. In general, it is not realistic to observe similar blowup patterns of a period of more than two days. A preliminary data analysis shows that blowups seem to appear in a short time period (2 or 3 days), while the transformed returns (negative, positive) have a much longer lag-k tail dependence.

Considering the properties of clustered data in two consecutive days and much longer lag-k tail dependencies in real data, the following parameter structure seems reasonable, i.e. we assume the matrix of weights (a_{lk}) to have the following structure:

$$(a_{lk}) = \begin{pmatrix} a_{00} & 0 & 0 & 0 & \cdots & 0 \\ a_{10} & a_{11} & 0 & 0 & \cdots & 0 \\ a_{20} & 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ a_{L0} & 0 & 0 & 0 & \dots & a_{LL} \end{pmatrix}. \tag{4.3}$$

Now the number L corresponds to the maximal lag of tail dependencies within the sequence; the lag-k tail dependence index is characterized by the coefficients a_{k0} and a_{kk}. The coefficient a_{00} represents the proportion of the number of observations which are drawn from an independent process $\{Z_{0i}\}$. In other words, a very large value at time 0 has no future impact when the large value is generated from $\{Z_{0i}\}$. If both a_{k0} and a_{kk} are not zero, then a very large value at time 0 has impact at time k when the large value is generated from $\{Z_{ki}\}$. If there is strong lag-k tail dependence for each k, the value of a_{00} will be small. While two coefficients a_{k0} and a_{kk} may not be sufficient enough to characterize different kinds of k step impacts from a very large value at time 0, the setups of (4.1) and (4.3) are considered as an approximation of the observed process.

It is clear that as long as the maximal lag of tail dependence of a sequence has been determined, we can easily write the model based on (4.1) and (4.3). Although the estimators for coefficients in a general M4 model have been proposed by Zhang and Smith (2003), Zhang (2003b), there would be advantages to use the special structure in (4.3) to construct estimators, i.e. the conditions imposed on the parameters can be reduced to a minimal level. We now compute the lag-r tail dependence index and then, in next section, turn to the estimation of the parameters in (4.3).
For $r > 0$, we have

$$\begin{aligned} &P(Y_1 \le x,\ Y_{1+r} \le y) \\ &\quad = P(a_{lk}Z_{l,1-k} \le x,\ a_{lk}Z_{l,1+r-k} \le y,\ 0 \le l \le L,\ 0 \le k \le L) \\ &\quad = P(a_{l0}Z_{l1} \le x,\ a_{ll}Z_{l,1-l} \le x,\ a_{l0}Z_{l,1+r} \le y,\ a_{ll}Z_{l,1+r-l} \le y,\ 0 \le l \le L) \\ &\quad = \exp\{-\sum_{l\neq r}(\frac{a_{l0}+a_{ll}}{x} + \frac{a_{r0}+a_{rr}}{y}) - \frac{a_{rr}}{x} - \frac{a_{r0}}{y} - \max(\frac{a_{r0}}{x},\ \frac{a_{rr}}{y})\} \\ &\quad = \exp\{-\frac{1}{x} - \frac{1}{y} + \min(\frac{a_{r0}}{x},\ \frac{a_{rr}}{y})\}. \end{aligned} \tag{4.4}$$

A general joint probability computation leads to the following expression:

$$\begin{aligned} P(Y_i \le y_i, 1 \le i \le r) &= P(Z_{l,i-k} \le \frac{y_i}{a_{l,k}} \text{ for } 0 \le l \le L,\ 0 \le k \le L,\ 1 \le i \le r) \\ &= P(Z_{l,m} \le \min_{1-m\le k\le r-m} \frac{y_{m+k}}{a_{l,k}}, 0 \le l \le L,\ -l+1 < m < r) \\ &= \exp\Big(-\sum_{l=0}^{L}\sum_{m=-l+1}^{r} \max_{1-m\le k\le r-m} \frac{a_{l,k}}{y_{m+k}}\Big). \end{aligned} \tag{4.5}$$

Since

$$\begin{aligned} \frac{P\big(Y_1 \ge u,\ Y_{1+r} \ge u\big)}{P\big(Y_1 \ge u\big)} &= 1 - \frac{\exp\{\frac{1}{u}\} - \exp\{-\frac{2-\min(a_{r0},\ a_{rr})}{u}\}}{1-\exp\{\frac{1}{u}\}} \\ &\to \min(a_{r0},\ a_{rr}) \end{aligned} \tag{4.6}$$

as $u \to \infty$, the lag-r tail dependence index of model (4.1) is $\min(a_{r0},\ a_{rr})$. There is some intuition behind the characterization of lag-r tail dependence index in (4.6). The value of $(a_{r0}+a_{rr})$ represents the proportion of the number of observations which are drawn from the process Z_{ri}. The value of $\min(a_{r0},\ a_{rr})$ represents the proportion of the number of observations which are over a certain threshold and are drawn from the lag-r dependence process. This can be seen from the left hand side of (4.6) when it is replaced by its empirical counterpart. The empirical counterpart can also be used to estimate $\min(a_{r0}, a_{rr})$ for a suitable choice of u value. Parameter estimation will be discussed in Section 6. We now illustrate an example and show how the gamma test detects the order of lag-k tail dependence.

Example 4.1 *Consider the model (4.1) with the following parameter structure:*

$$(a_{lk}) = \begin{pmatrix} 0.3765 & 0 & 0 & 0 & 0 & 0 \\ 0.0681 & 0.0725 & 0 & 0 & 0 & 0 \\ 0.0450 & 0 & 0.0544 & 0 & 0 & 0 \\ 0.0276 & 0 & 0 & 0.1166 & 0 & 0 \\ 0.0711 & 0 & 0 & 0 & 0.0185 & 0 \\ 0.1113 & 0 & 0 & 0 & 0 & 0.0386 \end{pmatrix}. \tag{4.7}$$

We use (4.1) and (4.7) to generate a sequence of observations of size 5000. These observations are plotted in Figure 6. We use the gamma test (2.12) to test lag-k tail dependencies at level $\alpha = .05$.

Figure 6 about here

As pointed out earlier, the index l in model (4.1) corresponds to a signature pattern of the observed process. The sum of $\sum_k a_{lk}$ yields the proportion of the total number of observations drawn from the lth independent sequences of Z_{lk}, $-\infty < k < \infty$. If we simply use all data for the gamma test, we may not get the right indications of the lag-k tail dependencies because the values computed from the test statistic may not be associated with the kth moving pattern. Here we conduct the gamma test in a local moving window of size 300. We randomly draw 100 local windows and perform the gamma test using the 300 observations in each window. The number of rejections of lag-k tail dependencies are summarized in the following table.

lag-k	1	2	3	4	5	6	7	8	9	10
# of rejection of H_0	7	5	2	3	5	1	1	0	0	0

The rejection rates of the lag-k, $k = 1, 2, 3, 4, 5$, are close to their corresponding proportions of observations which are drawn from the corresponding moving patterns. The rejection rates of larger lag-k are relatively small. Therefore the maximal lag of 5 for tail dependence present in the simulated data is indeed suggested by the gamma test results.

5 Combining M3 with a Markov process: a new nonlinear time series model

The previous analysis has suggested that there is asymmetric behavior between negative returns and positive returns. We have also tested whether there is tail dependence between positive returns and negative returns, and found that the null hypothesis of tail independence was not rejected. These phenomena suggest that models for negative returns should be different from models for positive returns.

First, we want to find a class of models of (4.1) and (4.3) to model negative returns. Second, we want to find a different class of models of (4.1) and (4.3) to model positive returns. Then we combine those two classes of models with a Markov process for both returns.

Notice that in the time series plot of negative returns, we have many zeros values (close to 50% of the total number of points) at which positive returns were observed. This suggests that models for negative returns should have a variable to model the locations of occurrences of zeros. This results in the following model structure.

Model 1: Combining M3 (used to model scales) with a Markov process (used to model signs): a new model for negative returns:

$$Y_i^- = \max_{0 \le l \le L^-} \max_{0 \le k \le K^-} a_{lk}^- Z_{l,i-k}^-, \quad -\infty < i < \infty,$$

where the superscript $^-$ means that the model is for negative returns only. Constants $\{a_{lk}^-\}$ are nonnegative and satisfy $\sum_{l=0}^{L^-} \sum_{k=0}^{K^-} a_{lk}^- = 1$. The matrix of weights is

$$(a_{lk}^-) = \begin{pmatrix} a_{00}^- & 0 & 0 & 0 & \cdots & 0 \\ a_{10}^- & a_{11}^- & 0 & 0 & \cdots & 0 \\ a_{20}^- & 0 & a_{22}^- & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ a_{L^-0}^- & 0 & 0 & 0 & \dots & a_{L^-L^-}^- \end{pmatrix}.$$

$\{Z_{li}^-,\ l = 1, \dots, L^-,\ -\infty < i < \infty\}$ is an independent array, where random variables Z_{li}^- are identically distributed with a unit Fréchet distribution function. Let

$$R_i^- = \xi_i^- Y_i^-, \ -\infty < i < \infty, \tag{5.1}$$

where the process $\{\xi_i^-\}$ is independent of $\{Y_i^-\}$ and takes values in a finite set $\{0,\, 1\}$ – i.e., $\{\xi_i^-\}$ is a sign process. Here $\{Y_i^-\}$ is an M3 process, $\{\xi_i^-\}$ is a simple Markov process. $\{R_i^-\}$ is the negative return process. For simplicity, Model (5.1) is regarded as MCM3 processes.

Remark 4 *If $\{Y_i^-\}$ is an independent process, then $P(R_{i+r}^- > u | R_i^- > u) \to 0$ as $u \to \infty$ for $i > 0$, $r > 0$, i.e. no tail dependence exists. This phenomenon tells that if there are tail dependencies in the observed process, the model with time dependence (through a Markov chain) only can not model the tail dependence if the random variables used to model scales are not tail dependent.*

Model 2: An MCM3 process model for positive returns:

$$Y_i^+ = \max_{0 \le l \le L^+} \max_{0 \le k \le K^+} a_{lk}^+ Z_{l,i-k}^+, \quad -\infty < i < \infty,$$

where the superscript $^+$ means that the model is for positive returns only. Constants $\{a_{lk}^+\}$ are nonnegative and satisfy $\sum_{l=0}^{L^+} \sum_{k=0}^{K^+} a_{lk}^+ = 1$. The matrix of weights is

$$(a_{lk}^+) = \begin{pmatrix} a_{00}^+ & 0 & 0 & 0 & \cdots & 0 \\ a_{10}^+ & a_{11}^+ & 0 & 0 & \cdots & 0 \\ a_{20}^+ & 0 & a_{22}^+ & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ a_{L^+0}^+ & 0 & 0 & 0 & \dots & a_{L^+L^+}^+ \end{pmatrix}.$$

Random variables $\{Z^{+}_{li},\ l = 1,\ldots,L^{+},\ -\infty < i < \infty\}$ is an independent array, where Z^{+}_{li}s are identically distributed with a unit Fréchet distribution. Let

$$R^{+}_{i} = \xi^{+}_{i} Y^{+}_{i},\ -\infty < i < \infty, \tag{5.2}$$

where the process $\{\xi^{+}_{i}\}$ is independent of $\{Y^{+}_{i}\}$ and takes values in a finite set $\{0, 1\}$. Here $\{Y^{+}_{i}\}$ is an M3 process, $\{\xi^{+}_{i}\}$ is a simple Markov process. $\{R^{+}_{i}\}$ is the positive return process.

Remark 5 *In previous sections, we have seen that negative returns Y^{-}_{i} and positive returns Y^{+}_{i} are asymmetric, and concluded that models for positive returns should be different from models for negative returns. Notice that at any time i, one can only observe one of the Y^{-}_{i} and Y^{+}_{i}. The other one is missing. By introducing the Markov processes ξ^{-}_{i} and ξ^{+}_{i}, both R^{-}_{i} in (5.1) and R^{+}_{i} in (5.2) are observable. We use R^{-}_{i} and R^{+}_{i} to construct parameter estimators.*

Model 3: An MCM3 process model for returns: with the established notations in (5.1) and (5.2), let

$$R_i = \mathrm{sign}(\xi_i) * [I_{\xi_i=-1} Y^{-}_{i} + I_{\xi_i=1} Y^{+}_{i}],\ -\infty < i < \infty, \tag{5.3}$$

where the process $\{\xi_i\}$ is a simple Markov process which is independent of $\{Y^{\pm}_{i}\}$ and takes values in a finite set {-1, 0, 1}. $\{R_i\}$ is the return process.

Remark 6 *The processes $\{\xi^{-}_{i}\}$, $\{\xi^{+}_{i}\}$ may be Bernoulli processes or Markov processes taking values in a finite set. The process $\{\xi_i\}$ may be considered as an independent process or a Markov process taking values in a finite set.*

Remark 7 *In Model (5.3), as long as Y^{-}_{i}, Y^{+}_{i}, and ξ_i are determined, R_i is determined.*

Remark 8 *In many applications, only positive observed values are concerned. Insurance claims, annual maxima of precipitations, file sizes, durations in internet traffic at certain point are some of those examples having positive values only. Even in our negative return model, the values have been converted into positive values. Therefore, the properties of Model (5.1) can easily extended to Model (5.2). We now focus on Model (5.1).*

We first compute the lag-k tail dependence index in Model (5.1):

$$\begin{aligned}
\frac{P(R^{-}_{j} > u, R^{-}_{j+k} > u)}{P(R^{-}_{j} > u)} &= \frac{P(Y^{-}_{j} > u, \xi^{-}_{j} = 1, Y^{-}_{j+k} > u, \xi^{-}_{j+k} = 1)}{P(Y^{-}_{j} > u, \xi^{-}_{j} = 1)} \\
&= \frac{P(Y^{-}_{j} > u, Y^{-}_{j+k} > u)}{P(Y^{-}_{j} > u)} \frac{P(\xi^{-}_{j} = 1, \xi^{-}_{j+k} = 1)}{P(\xi^{-}_{j} = 1)} \\
&= \frac{P(Y^{-}_{j} > u, Y^{-}_{j+k} > u)}{P(Y^{-}_{j} > u)} P(\xi^{-}_{j+k} = 1 | \xi^{-}_{j} = 1) \\
&\to \min(a^{-}_{k0}, a^{-}_{kk}) P(\xi^{-}_{j+k} = 1 | \xi^{-}_{j} = 1),
\end{aligned} \tag{5.4}$$

as u tends to infinite.

Now suppose $\{\xi_i^-\}$ is a simple Markov process taking values in a finite set $\{0,\ 1\}$, and with the transition probabilities:

$$\begin{aligned} &P(\xi_{j+1}^- = 0|\xi_j^- = 0) = p_{00},\ P(\xi_{j+1}^- = 1|\xi_j^- = 0) = p_{01}, \\ &P(\xi_{j+1}^- = 0|\xi_j^- = 1) = p_{10},\ P(\xi_{j+1}^- = 1|\xi_j^- = 1) = p_{11}, \end{aligned} \tag{5.5}$$

and the kth $(k > 1)$ step transition probabilities:

$$\begin{aligned} &P(\xi_{j+k}^- = 0|\xi_j^- = 0) = p_{00}^{(k)},\ P(\xi_{j+k}^- = 1|\xi_j^- = 0) = p_{01}^{(k)}, \\ &P(\xi_{j+k}^- = 0|\xi_j^- = 1) = p_{10}^{(k)},\ P(\xi_{j+k}^- = 1|\xi_j^- = 1) = p_{11}^{(k)}, \end{aligned} \tag{5.6}$$

where the superscripts (k) denote the kth step in a Markov process.
Using (5.5) and (5.6), we first have

$$\begin{aligned} P(R_1^- < x, R_{1+r}^- < y) &= P(R_1^- < x, R_{1+r}^- < y, \xi_1^- = 0) + P(R_1^- < x, R_{1+r}^- < y, \xi_1^- = 1) \\ &= P(R_{1+r}^- < y,\ \xi_1^- = 0) + P(Y_1^- < x, R_{1+r}^- < y,\ \xi_1^- = 1). \end{aligned} \tag{5.7}$$

Since

$$\begin{aligned} P(R_{1+r}^- < y,\ \xi_1^- = 0) &= P(R_{1+r}^- < y, \xi_{1+r}^- = 0,\ \xi_1^- = 0) + P(R_{1+r}^- < y, \xi_{1+r}^- = 1,\ \xi_1^- = 0) \\ &= P(\xi_{1+r}^- = 0,\ \xi_1^- = 0) + P(Y_{1+r}^- < y,\ \xi_{1+r}^- = 1,\ \xi_1^- = 0) \\ &= P(\xi_1^- = 0)P(\xi_{1+r}^- = 0|\xi_1^- = 0) \\ &\qquad + P(Y_{1+r}^- < y)P(\xi_1^- = 0)P(\xi_{1+r}^- = 1|\xi_1^- = 0) \\ &= p_0 p_{00}^{(r)} + p_0 p_{01}^{(r)} e^{-1/y}, \end{aligned} \tag{5.8}$$

where $p_0 = P(\xi_1^- = 0)$ is the probability of the chain starting at the initial state $\xi_1^- = 0$;

$$\begin{aligned} P(Y_1^- < x, R_{1+r}^- < y, \xi_1^- = 1) &= P(Y_1^- < x, R_{1+r}^- < y, \xi_{1+r}^- = 0, \xi_1^- = 1) \\ &\qquad + P(Y_1^- < x, R_{1+r}^- < y, \xi_{1+r}^- = 1, \xi_1^- = 1) \\ &= P(Y_1^- < x, \xi_{1+r}^- = 0, \xi_1^- = 1) \\ &\qquad + P(Y_1^- < x, Y_{1+r}^- < y)P(\xi_{1+r}^- = 1, \xi_1^- = 1) \\ &= p_1 p_{10}^{(r)} e^{-1/x} + p_1 p_{11}^{(r)} \exp\{-\frac{1}{x} - \frac{1}{y} + \min(\frac{a_{r0}^-}{x}, \frac{a_{rr}^-}{y})\}, \end{aligned} \tag{5.9}$$

where $p_1 = 1 - p_0$; then putting (5.8) and (5.9) in (5.7), we have

$$P(R_1^- < x, R_{1+r}^- < y) = p_0 p_{00}^{(r)} + p_0 p_{01}^{(r)} e^{-1/y} + p_1 p_{10}^{(r)} e^{-1/x} + p_1 p_{11}^{(r)} \exp\{-\frac{1}{x} - \frac{1}{y} + \min(\frac{a_{r0}^-}{x}, \frac{a_{rr}^-}{y})\}. \tag{5.10}$$

Notice that the event $\{R_i^- < x\}$ is a union of two events $\{Y_i^- < x,\ \xi_i^- = 1\}$ and $\{\xi_i^- = 0\}$, which are mutually exclusive. Then for $i_1 < i_2 < \cdots < i_m$, we have the following joint probability expression:

$$\begin{aligned}
&P\big(R^-_{i_1} < x_{i_1},\ R^-_{i_2} < x_{i_2}, \ldots, R^-_{i_m} < x_{i_m}\big) \\
&= p_1 \prod_{j=1}^{m-1} p_{11}^{(i_{j+1}-i_j)} P\big(Y^-_{i_1} < x_{i_1},\ Y^-_{i_2} < x_{i_2}, \ldots, Y^-_{i_m} < x_{i_m}\big) \\
&\quad + \sum_{k=1}^{m-1} \Big[\sum_{\substack{j_1<j_2<\cdots<j_k \\ \{j_1,j_2,\ldots,j_k\}\subset\{i_1,i_2,\ldots,i_m\}}} p_0^{I_{(j_1>i_1)}} p_1^{I_{(j_1=i_1)}} \prod_{j=1}^{m-1} p^{(i_{j+1}-i_j)}_{I_{(i_j\in\{j_1,j_2,\ldots,j_k\})}I_{(i_{j+1}\in\{j_1,j_2,\ldots,j_k\})}} \qquad (5.11)\\
&\qquad P\big(Y^-_{j_1} < x_{j_1},\ Y^-_{j_2} < x_{j_2}, \ldots, Y^-_{j_k} < x_{j_k}\big)\Big] \\
&\quad + p_0 \prod_{j=1}^{m-1} p_{00}^{(i_{j+1}-i_j)}.
\end{aligned}$$

Notice that an exact M4 data generating process (DGP) may not be observable in real – i.e., it may not be realistic to observe an infinite number of times of signature patterns as demonstrated in Figure 5. It is natural to consider the following model:

$$R^*_i = \xi^-_i (Y^-_i + N^-_i),\ -\infty < i < \infty, \qquad (5.12)$$

where $\{N^-_i\}$ is an independent bounded noise process which is also independent of $\{\xi^-_i\}$ and $\{Y^-_i\}$. By adding the noise process, the signature patterns can not be explicitly illustrated as we did in Figure 5.

The following proposition tells that as long as the characterization of tail dependencies is the main concern, Model (5.1) or its simplified form should be a good approximation to the possible true model.

Proposition 5.1 *The lag-k tail dependence within $\{R^*_i\}$ can be expressed as:*

$$\frac{P(R^*_j > u, R^*_{j+k} > u)}{P(R^*_j > u)} \to \min(a^-_{k0}, a^-_{kk}) P(\xi^-_{j+k} = 1 | \xi^-_j = 1) \qquad (5.13)$$

as $u \to \infty$.

A proof of Proposition 5.1 is given in Section 9.

The same lag-k tail dependence index in both (5.4) and (5.13) suggests that an MCM3 process (5.1) can be used to approximate (5.12). Under (5.12), it is reasonable to assume each paired parameters being identical. Under this assumption, we derive the parameter estimators in the next section.

6 MCM3 Parameter estimation

Notice that on the right hand sides of (5.4) and (5.13) are moving coefficients and the rth step transition probabilities. We choose to substitute the quantities on the left hand sides by their corresponding empirical counterparts to construct the parameter estimators.

Considering that $\{\xi_i^-\}$ and $\{Y_i^-\}$ are assumed independent, the estimations of parameters from these two processes can be processed separately. Our main interest here is to estimate the parameters in M3 process. The maximum likelihood estimations and the asymptotic normality of the estimators for Markov processes have been developed in Billingsley (1961a,b). Here we simply perform empirical estimation and then treat the estimated transition probabilities as known parameter values to accomplish statistical inference for the M3 processes.

We now assume $a_{r0} = a_{rr}$ for all $r = 1, \ldots, L = L^-$, and denote $P(R_i^- > u,\ R_{i+r}^- > u)$ as μ_r^- for $r = 1, \ldots, L$, $P(R_1^- > u)$ as μ_{L+1}^-, a_{r0} as a_r respectively.
Define

$$\bar{X}_r^- = \frac{1}{n}\sum_{i=1}^{n-r} I_{(R_i^- > u,\ R_{i+r}^- > u)},\ r = 1, \ldots, L, \tag{6.1}$$

$$\bar{X}_{L+1}^- = \frac{1}{n}\sum_{i=1}^{n-r} I_{(R_i^- > u)}. \tag{6.2}$$

Then by the strong law of large numbers (SLLN), we have

$$\bar{X}_r^- \xrightarrow{a.s.} P(R_i^- > u,\ R_{i+r}^- > u) = \mu_r^-,\ r = 1, \ldots, L, \tag{6.3}$$

$$\bar{X}_{L+1} \xrightarrow{a.s.} P(R_1^- > u) = \mu_{L+1}^-. \tag{6.4}$$

From (5.4) and (5.13), we propose the following estimators for parameters a_r^-:

$$\widehat{a}_r^- = \frac{\bar{X}_r^-}{\bar{X}_{L+1}^- p_{11}^{(r)}},\ r = 1, ..., L. \tag{6.5}$$

In order to study asymptotic normality, we introduce the following proposition which is Theorem 27.4 in Billingsley (1995). First we introduce the so-called α-mixing condition.

For a sequence $Y_1, Y_2, ...$ of random variables, let α_n be a number such that

$$\mid P(A \cap B) - P(A)P(B) \mid \leq \alpha_n$$

for $A \in \sigma(Y_1, ..., Y_k)$, $B \in \sigma(Y_{k+n}, Y_{k+n+1}, ...)$, and $k \geq 1, n \geq 1$. When $\alpha_n \to 0$, the sequence $\{Y_n\}$ is said to be α-mixing.

Proposition 6.1 *Suppose that $X_1, X_2, \ldots$ is stationary and α-mixing with $\alpha_n = O(n^{-5})$ and that $E[X_n] = 0$ and $E[X_n^{12}] < \infty$. If $S_n = X_1 + \cdots + X_n$, then*

$$n^{-1}\, Var[S_n] \to \sigma^2 = E[X_1^2] + 2\sum_{k=1}^{\infty} E[X_1 X_{1+k}],$$

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

Remark 9 *The conditions $\alpha_n = O(n^{-5})$ and $E[X_n^{12}] < \infty$ are stronger than necessary as stated in the remark following Theorem 27.4 in Billingsley (1995) to avoid technical complication in the proof.*

With the established notations, we have the following lemma dealing with the asymptotic properties of the empirical functions.

Lemma 6.2 *Suppose that $\bar{X}_j^-$ and μ_j^- are defined in (6.1)-(6.4), then*

$$\sqrt{n}\left(\begin{bmatrix}\bar{X}_1^-\\ \vdots\\ \bar{X}_{L+1}^-\end{bmatrix}-\begin{bmatrix}\mu_1^-\\ \vdots\\ \mu_{L+1}^-\end{bmatrix}\right)\xrightarrow{\mathcal{L}} N\Big(0,\Sigma^-+\sum_{k=1}^{L}\{W_k^-+W_k^{-T}\}\Big),$$

where the entries σ_{ij} of matrix Σ^- and the entries w_k^{ij} of the matrix W_k^- are defined below. For $i=1,\dots,L,\ j=1,\dots,L,$

$$\begin{aligned}\sigma_{ij}=&E(I_{\{R_1^->u,\ R_{1+i}^->u\}}-\mu_i^-)(I_{\{R_1^->u,\ R_{1+j}^->u\}}-\mu_j^-)\\ =&P(R_1^->u,\ R_{1+i}^->u,\ R_{1+j}^->u)-\mu_i^-\mu_j^-.\end{aligned}$$

$$\begin{aligned}w_k^{ij}=&E(I_{\{R_1^->u,\ R_{1+i}^->u\}}-\mu_i^-)(I_{\{R_{1+k}^->u,\ R_{1+j+k}^->u\}}-\mu_j^-)\\ =&P(R_1^->u,\ R_{1+i}^->u,R_{1+k}^->u,\ R_{1+j+k}^->u)-\mu_i^-\mu_j^-.\end{aligned}$$

For $i=1,\dots,L,\ j=L+1,$

$$\begin{aligned}\sigma_{ij}=&E(I_{\{R_1^->u,\ R_{1+i}^->u\}}-\mu_i^-)(I_{\{R_1^->u\}}-\mu_j^-)\\ =&\mu_i^--\mu_i^-\mu_j^-.\end{aligned}$$

$$\begin{aligned}w_k^{ij}=&E(I_{\{R_1^->u,\ R_{1+i}^->u\}}-\mu_i^-)(I_{\{R_{1+k}^->u\}}-\mu_j^-)\\ =&P(R_1^->u,\ R_{1+i}^->u,R_{1+k}^->u)-\mu_i^-\mu_j^-.\end{aligned}$$

For $i=L+1,\ j=1,\dots,L,$

$$\sigma_{ij}=\sigma_{ji}.$$

$$\begin{aligned}w_k^{ij}=&E(I_{\{R_1^->u\}}-\mu_i^-)(I_{\{R_{1+k}^->u,\ R_{1+j+k}^->u\}}-\mu_j^-)\\ =&P(R_1^->u,R_{1+k}^->u,\ R_{1+j+k}^->u)-\mu_i^-\mu_j^-.\end{aligned}$$

For $i=L+1,\ j=L+1,$

$$\sigma_{ij}=\mu_i^--\mu_i^-\mu_j^-.$$

$$\begin{aligned}w_k^{ij}=&E(I_{\{R_1^->u\}}-\mu_i^-)(I_{\{R_{1+k}^->u\}}-\mu_j^-)\\ =&P(R_1^->u,\ R_{1+k}^->u)-\mu_i^-\mu_j^-=\mu_k^--\mu_i^-\mu_j^-.\end{aligned}$$

A proof of Lemma 6.2 is deferred to Section 9.

From (6.5), we have the following Jacobian matrix

$$\Theta^-=\begin{pmatrix}\frac{1}{\mu_{L+1}^-p_{11}} & \cdots & \cdots & \cdots & \cdots & -\frac{\mu_1^-}{(\mu_{L+1}^-)^2p_{11}}\\ & \frac{1}{\mu_{L+1}^-p_{11}^{(2)}} & & & & -\frac{\mu_2^-}{(\mu_{L+1}^-)^2p_{11}^{(2)}}\\ & & & \ddots & & \\ & & & & \frac{1}{\mu_{L+1}^-p_{11}^{(L)}} & -\frac{\mu_L^-}{(\mu_{L+1}^-)^2p_{11}^{(L)}}\end{pmatrix} \qquad (6.6)$$

which is applied to getting asymptotic covariance matrix for parameter estimators.

We have obtained the following theorem.

Theorem 6.3 *Let* $\widehat{a}_0^- = 1 - 2\sum_{i=1}^{L} \widehat{a}_i^-$*. Let* $\widehat{\mathbf{a}}^- = [\widehat{a}_0^-, \widehat{a}_1^-, \cdots, \widehat{a}_L^-]^T$*,* $\mathbf{a}^- = [a_0^-, a_1^-, \cdots, a_L^-]^T$ *be two vectors. Then with the established notations,*

$$\sqrt{n}(\widehat{\mathbf{a}}^- - \mathbf{a}^-) \xrightarrow{\mathcal{L}} N(0, C^- B^- C^{-T}),$$

where B^- *is a matrix with elements* $B_{1j}^- = 0,\ B_{i1}^- = 0,\ i, j = 1, 2, \ldots, m$*, and the minor of* B_{11}^- *being* $\Theta^-[\Sigma^- + \sum_{k=1}^{L}\{W_k^- + W_k^{-T}\}]\Theta^{-T}$*; and* C^- *is a matrix with elements* $C_{11}^- = 1,\ C_{1j}^- = -2,\ C_{i1}^- = 0,\ i, j = 2, \ldots, m$*, and the minor of* C_{11}^- *being a unit matrix.*

Similarly, we can construct estimators of parameters in Model 2. For Model 3, we propose to apply Model 1 and Model 2 first, and then re-estimate the one step Markov transition probabilities. The asymptotic properties of the transition probability estimators can be found in Billingsley (1961a,b). We also can derive the joint asymptotic covariance matrix for all parameter estimators following the similar process used for the combined M3 and Markov process for negative returns. We will not pursue that in this paper because we think the joint asymptotic properties from Model 1 and Model 2 may be enough for various practical purposes.

7 Modeling jumps in returns

7.1 Modeling jumps in negative returns

The previous analysis in Section 3 has suggested up to lag-12 tail dependencies for the negative. We now fit the model (5.1) to the transformed negative returns.

From the data, the proportion of the days that the negative returns are observed, i.e. the investors lose money, is 0.4731. We use Markov chain to model negative signs. Suppose the state 1 corresponds to the day that a negative return is observed and the state 0 corresponds to the day that a negative return is not observed. The one step transition probabilities $P(\xi_d = i|\xi_d = j),\ i, j = 0, 1$, are estimated in the following table.

State	0	1
0	0.5669	0.4331
1	0.4823	0.5177

They are empirical estimations – for example, $P(\xi_{i+1,d} = 0|\xi_{id} = 0) = P(Y_{i+1,d} \leq 0|Y_{i,d} \leq 0)$ is estimated by: $\sum_{i=1}^{n-1} I_{(Y_{i,d}\leq 0, Y_{i+1,d}\leq 0)} / \sum_{i=1}^{n-1} I_{(Y_{i,d}\leq 0)}$. The following table estimates the rth step transition probabilities using the data and computes the transition probabilities using Chapman-Kolmogorov equation.

Step	Estimated				Chapman-Kolmogorov			
1	0.5669	0.4331	0.4823	0.5177	0.5669	0.4331	0.4823	0.5177
2	0.5172	0.4828	0.5377	0.4623	0.5303	0.4697	0.5231	0.4769
3	0.5195	0.4805	0.5353	0.4647	0.5272	0.4728	0.5266	0.4734
4	0.5302	0.4698	0.5232	0.4768	0.5269	0.4731	0.5269	0.4731
5	0.5284	0.4716	0.5251	0.4749	0.5269	0.4731	0.5269	0.4731
6	0.5223	0.4777	0.5319	0.4681	0.5269	0.4731	0.5269	0.4731
7	0.5279	0.4721	0.5256	0.4744	0.5269	0.4731	0.5269	0.4731
8	0.5282	0.4718	0.5256	0.4744	0.5269	0.4731	0.5269	0.4731
9	0.5239	0.4761	0.5301	0.4699	0.5269	0.4731	0.5269	0.4731
10	0.5285	0.4715	0.5252	0.4748	0.5269	0.4731	0.5269	0.4731
11	0.5270	0.4730	0.5271	0.4729	0.5269	0.4731	0.5269	0.4731
12	0.5369	0.4631	0.5158	0.4842	0.5269	0.4731	0.5269	0.4731

One can see from the table that the estimated values (the middle panel) are very close to the theoretical values (the right panel) after the first step. The limiting distribution of the two state Markov chain is consistent with the proportions of the days that a negative return is observed. Based on this table, one can conclude that the data suggests that a two state Markov chain model is a good fit of the transitions of signs of negative returns. This analysis together with the previous analysis suggest that an MCM3 model is suitable for jumps in returns. Next we estimate the parameters in M3 model. The results are summarized in Table 3. The estimated parameter values can be used to further statistical inference – for example, to compute value at risk (VaR) based on the estimated model.

Table 3 about here

7.2 Modeling jumps in positive returns

Similar to the case of negative returns, the estimated first order transition probability matrix for positive returns is summarized in the following table.

State	0	1
0	0.5209	0.4791
1	0.4379	0.5621

The following table estimates the rth step transition probabilities using the data and computes the transition probabilities using Chapman-Kolmogorov equation.

Step	Estimated				Chapman-Kolmogorov			
1	0.5209	0.4791	0.4379	0.5621	0.5209	0.4791	0.4379	0.5621
2	0.4649	0.5351	0.4893	0.5107	0.4811	0.5189	0.4742	0.5258
3	0.4665	0.5335	0.4876	0.5124	0.4778	0.5222	0.4773	0.5227
4	0.4804	0.5196	0.4749	0.5251	0.4776	0.5224	0.4775	0.5225
5	0.4786	0.5214	0.4766	0.5234	0.4775	0.5225	0.4775	0.5225
6	0.4722	0.5278	0.4825	0.5175	0.4775	0.5225	0.4775	0.5225
7	0.4782	0.5218	0.4772	0.5228	0.4775	0.5225	0.4775	0.5225
8	0.4791	0.5209	0.4762	0.5238	0.4775	0.5225	0.4775	0.5225

The estimated parameter values in M3 model are summarized in Table 4. The estimated parameter values can be used to further statistical inference – for example, to compute value at risk (VaR) based on the estimated model. However, we restrict ourself to find estimates of the M3 process in this current work.

Table 4 about here

8 Discussion

In this paper, we obtained statistical evidences of the existence of extreme impacts in financial time series data; we introduced a new time series model structure – i.e., combinations of Markov processes, GARCH(1,1) volatility model, and M3 processes. We restricted our attentions to a subclass of M3 processes. This subclass has advantages of efficiently modeling serial tail dependent financial time series, while, of course, other model specifications are possibly also suitable.

We proposed models for tail dependencies in jumps in returns. The next step is to make statistical and economic inferences of computing risk measures and constructing prediction intervals. In a different project, we extend the results developed in this paper to study extremal risk analysis and portfolio choice.

The approach adopted in the paper is a hierarchical model structure – i.e., to apply GARCH(1,1) fitting and to get estimated standard deviations first; then based on standardized return series, we apply M3 and Markov processes modelling. It is possible to study Markov processes, GARCH processes, and M3 processes simultaneously. But that requires additional work. We put this goal as a direction for future research.

9 Appendix

Proof of Proposition 5.1. We have

$$\begin{aligned} P(R_j^* > u,\ R_{j+k}^* > u) &= P\big(Y_j^- + N_j^- > u,\ Y_{j+k}^- + N_{j+k}^- > u,\ \xi_j^- = 1,\ \xi_{j+k}^- = 1\big) \\ &= p_1 p_{11}^{(k)} P\big(Y_j^- + N_j^- > u,\ Y_{j+k}^- + N_{j+k}^- > u\big), \\ P(R_j^* > u) &= p_1 P\big(Y_j^- + N_j^- > u\big). \end{aligned}$$

Let $f(x)$ and M be the density and the bound limit of N_j, then

$$\frac{P(R_j^* > u,\ R_{j+k}^* > u)}{P(R_j^* > u)} = p_{11}^{(k)} \frac{P\big(Y_j^- + N_j^- > u,\ Y_{j+k}^- + N_{j+k}^- > u\big)}{P\big(Y_j^- + N_j^- > u\big)}$$
$$= p_{11}^{(k)} \frac{\int_{-M}^{M}\int_{-M}^{M} P\big(Y_j^- > u-x,\ Y_{j+k}^- > u-y\big) f(x)f(y)dxdy}{\int_{-M}^{M} P\big(Y_j^- > u-x\big) f(x)dx}$$
$$= p_{11}^{(k)} \frac{\int_{-M}^{M}\int_{-M}^{M} \frac{P\big(Y_j^- > u-x,\ Y_{j+k}^- > u-y\big)}{P\big(Y_j^- > u\big)} f(x)f(y)dxdy}{\int_{-M}^{M} \frac{P\big(Y_j^- > u-x\big)}{P\big(Y_j^- > u\big)} f(x)dx}.$$

It is easy to see that $\lim_{u\to\infty} \frac{P\big(Y_j^- > u-x\big)}{P\big(Y_j^- > u\big)} = 1$, and

$$\frac{P\big(Y_j^- > u-x,\ Y_{j+k}^- > u-y\big)}{P\big(Y_j^- > u\big)} = \frac{P\big(Y_j^- > u-x\big)}{P\big(Y_j^- > u\big)} \frac{P\big(Y_j^- > u-x,\ Y_{j+k}^- > u-y\big)}{P\big(Y_j^- > u-x\big)}.$$

We have

$$\frac{P(Y_j^- > w,\ Y_{j+k}^- > w+z)}{P(Y_j^- > w)} = \frac{1 - e^{-1/w} - e^{-1/(w+z)} + e^{-1/w-1/(w+z)+\min(a_{k0}/w, a_{kk}/(w+z))}}{1-e^{-1/w}}$$
$$= 1 - e^{-1/(w+z)}\big[\frac{1 - e^{-1/w+\min(a_{k0}/w, a_{kk}/(w+z))}}{1-e^{-1/w}}\big].$$

Since for $w > 0,\ z > 0$ (similarly for $z < 0$), we have

$$\min(a_{k0}/(w+z), a_{kk}/(w+z)) \le \min(a_{k0}/w, a_{kk}/(w+z)) \le \min(a_{k0}/w, a_{kk}/w),$$

so

$$\frac{1 - e^{-1/w+\min(a_{k0},a_{kk})/w}}{1-e^{-1/w}} \le \frac{1 - e^{-1/w+\min(a_{k0}/w, a_{kk}/(w+z))}}{1-e^{-1/w}} \le \frac{1 - e^{-1/w+\min(a_{k0},a_{kk})/(w+z)}}{1-e^{-1/w}}$$

which gives

$$\lim_{w\to\infty} \frac{1 - e^{-1/w+\min(a_{k0},a_{kk})/w}}{1-e^{-1/w}} = 1 - \min(a_{k0}, a_{kk}),$$
$$\lim_{w\to\infty} \frac{1 - e^{-1/w+\min(a_{k0},a_{kk})/(w+z)}}{1-e^{-1/w}} = 1 - \min(a_{k0}, a_{kk}),$$

hence

$$\lim_{w\to\infty} \frac{1 - e^{-1/w+\min(a_{k0}/w, a_{kk}/(w+z))}}{1-e^{-1/w}} = 1 - \min(a_{k0}, a_{kk}).$$

So we have

$$\lim_{w\to\infty} \frac{P(Y_j^- > w,\ Y_{j+k}^- > w+z)}{P(Y_j^- > w)} = \min(a_{k0}, a_{kk}).$$

Let $w = u - x$, $z = x - y$, then

$$\lim_{u\to\infty} \frac{P(Y_j^- > u-x,\ Y_{j+k}^- > u-y)}{P(Y_j^- > u)} = \min(a_{k0}, a_{kk})$$

which gives the desired results in (5.13). □

Proof of Lemma 6.2. Let

$$U_1 = (I_{(R_1^- > u, R_2^- > u)} - \mu_1^-, I_{(R_1^- > u, R_3^- > u)} - \mu_2^-, \ldots, I_{(R_1^- > u, R_{1+L}^- > u)} - \mu_L^-, I_{(R_1^- > u)} - \mu_{1+L}^-)^T,$$

$$U_{1+k} = (I_{(R_{1+k}^- > u, R_{2+k}^- > u)} - \mu_1^-, I_{(R_{1+k}^- > u, R_{3+k}^- > u)} - \mu_2^-, \ldots, I_{(R_{1+k}^- > u, R_{1+k+L}^- > u)} - \mu_L^-, I_{(R_{1+k}^- > u)} - \mu_{1+L}^-)^T,$$

and $\alpha = (\alpha_1, \ldots, \alpha_L)^T \neq 0$ be an arbitrary vector.

Let $X_1 = \alpha^T U_1, X_{1+k} = \alpha^T U_{1+k}, \ldots$, then $E[X_n] = 0$ and $E[X_n^{12}] < \infty$. So Proposition 6.1 can apply. We say expectation are applied on all elements if expectation is applied on a random matrix. But $E[X_1^2] = \alpha^T E[U_1 U_1^T]\alpha = \alpha^T \Sigma \alpha$, $E[X_1 X_{1+k}] = \alpha^T E[U_1 U_{1+k}^T]\alpha = \alpha^T W_k \alpha$. The entries of Σ and W_k are computed from the following expressions.

For $i = 1, \ldots, L,\ j = 1, \ldots, L$,

$$\begin{aligned}\sigma_{ij} =& E(I_{\{R_1^- > u,\ R_{1+i}^- > u\}} - \mu_i^-)(I_{\{R_1^- > u,\ R_{1+j}^- > u\}} - \mu_j^-)\\ =& P(R_1^- > u,\ R_{1+i}^- > u,\ R_{1+j}^- > u) - \mu_i^- \mu_j^-.\end{aligned}$$

$$\begin{aligned}w_k^{ij} =& E(I_{\{R_1^- > u,\ R_{1+i}^- > u\}} - \mu_i^-)(I_{\{R_{1+k}^- > u,\ R_{1+j+k}^- > u\}} - \mu_j^-)\\ =& P(R_1^- > u,\ R_{1+i}^- > u, R_{1+k}^- > u,\ R_{1+j+k}^- > u) - \mu_i^- \mu_j^-.\end{aligned}$$

For $i = 1, \ldots, L,\ j = L + 1$,

$$\begin{aligned}\sigma_{ij} =& E(I_{\{R_1^- > u,\ R_{1+i}^- > u\}} - \mu_i^-)(I_{\{R_1^- > u\}} - \mu_j^-)\\ =& \mu_i^- - \mu_i^- \mu_j^-.\end{aligned}$$

$$\begin{aligned}w_k^{ij} =& E(I_{\{R_1^- > u,\ R_{1+i}^- > u\}} - \mu_i^-)(I_{\{R_{1+k}^- > u\}} - \mu_j^-)\\ =& P(R_1^- > u,\ R_{1+i}^- > u, R_{1+k}^- > u) - \mu_i^- \mu_j^-.\end{aligned}$$

For $i = L + 1,\ j = 1, \ldots, L$,

$$\sigma_{ij} = \sigma_{ji}.$$

$$\begin{aligned}w_k^{ij} =& E(I_{\{R_1^- > u\}} - \mu_i^-)(I_{\{R_{1+k}^- > u,\ R_{1+j+k}^- > u\}} - \mu_j^-)\\ =& P(R_1^- > u, R_{1+k}^- > u,\ R_{1+j+k}^- > u) - \mu_i^- \mu_j^-.\end{aligned}$$

For $i = L + 1,\ j = L + 1$,

$$\sigma_{ij} = \mu_i^- - \mu_i^- \mu_j^-.$$

$$\begin{aligned}w_k^{ij} =& E(I_{\{R_1^- > u\}} - \mu_i^-)(I_{\{R_{1+k}^- > u\}} - \mu_j^-)\\ =& P(R_1^- > u,\ R_{1+k}^- > u) - \mu_i^- \mu_j^- = \mu_k^- - \mu_i^- \mu_j^-.\end{aligned}$$

So the proof is completed by applying the Cramér-Wold device. □

Acknowledgement

This work was supported in part by NSF Grants DMS-0443048 and DMS-0505528, and Institute for Mathematical Research, ETH Zürich. The author thanks Professor Paul Embrechts, Professor Tom Fomby (Editor), and one referee for their many comments, suggestions which have greatly improved the quality of the paper.

References

[1] Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control ac-19, 716–723.

[2] Billingsley, P., 1961a. Statistical Inference for Markov Processes . University of Chicago Press.

[3] Billingsley, P., 1961b. Statistical methods in Markov chains. Annals of Mathematical Statistics, 32, 12-40.

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

[5] Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

[6] Danielsson, J., 2002. The emperor has no clothes: limits to risk modelling. Journal of Banking and Finance 26, 1273–1296.

[7] Davis, R. A., and Resnick, S. I., 1989. Basic properties and prediction of max-ARMA processes. Advances in Applied Probability , 21, 781–803.

[8] Davis, R. A. and Resnick, S. I., 1993. Prediction of stationary max-stable processes. Annals of Applied Probability 3, 497–525.

[9] Deheuvels, P., 1983. Point processes and multivariate extreme values. Journal of Multivariate Analysis 13, 257–272.

[10] Duan, J. C., Ritchken, P., and Sun, Z., 2003. Option valuation with jumps in returns and volatility. Web reference.

[11] Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997 Modelling Extremal Events for Insurance and Finance. Springer

[12] Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls, in: M.A.H. Dempster (Eds.) Risk Management: Value at Risk and Beyond. Cambridge University Press, Cambridge, pp. 176–223

[13] Engle, R., 2002. Dynamic conditional correlation - a simple class of multivariate GARCH models. Journal of Business and Economic Statistics 17,339-50.

[14] Eraker, B., Johannes, M, and Polson, N. G., 2003. The impact of jumps in returns and volatility. Journal of Finance 53, 1269–1300.

[15] de Haan, L., 1984. A spectral representation for max-stable processes. Annals of Probability 12, 1194–1204.

[16] de Haan, L. and Resnick, S. I., 1977. Limit theory for multivariate sample extremes. Z. Wahrscheinlichkeitstheorie verw. Gebiete 40, 317–337.

[17] 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.

[18] Hall, P. and Yao. Q., 2002. Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica 71, 285–317.

[19] Leadbetter, M. R., Lindgren, G. and Rootzén, H., 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.

[20] McNeil, A. and Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7, 271–300.

[21] Mikosch, T. and Stărică, C., 2003. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Annals of Statististics 28, 1427–1451.

[22] Mikosch, T. and Straumann, D., 2002. Whittle estimation in a heavy-tailed GARCH(1,1) model. Stochastic Process. Appl,, 100: 187–222.

[23] Nandagopalan, S., 1990. Multivariate Extremes and the Estimation of the Extremal Index. Ph.D. dissertation, Dept. Of Statistics, University of North Carolina, Chapel Hill.

[24] Nandagopalan, S., 1994. On the multivariate extremal index. Journal of Research, National Institute of Standards and Technology 99, 543–550.

[25] Pickands, J. III, 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3, 119–131.

[26] Resnick, S. I., 1987. Extreme Values, Regular Variation, and Point Processes. Springer, New York.

[27] Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

[28] Sibuya, M., 1960. Bivariate extreme statistics, I. Annals of Institute of Statistical Mathematics 1, 195–210.

[29] Smith, R.L., 2003. Statistics of extremes, with applications in the environment, insurance and finance, in: B. Finkenstadt and H. Rootzén, (Eds.), Extreme Value in Finance, Telecommunications and the Environment. Chapman and Hall/CRC, pp. 1–78.

[30] Smith, R. L., and Weissman, I., 1996. Characterization and estimation of the multivariate extremal index. Technical report, the University of North Carolina.

[31] Straumann, D., 2003. Estimation in Conditionally Heteroscedastic Time Series Models. PhD thesis. Laboratory of Actuary Mathematics, University of Copenhagen.

[32] Tsay, R. S., 1999. Extreme value analysis of financial data. Manuscript, Univ. Of Chicago.

[33] Zhang, Z., 2003a. Quotient correlation: a sample based alternative to Pearson's correlation. Manuscript, Washington University.

[34] Zhang, Z., 2003b. The estimation of M4 processes with geometric moving patterns. Manuscript, Washington University.

[35] Zhang, Z., 2004. Some results on approximating max-stable processes. Manuscript, Washington University.

[36] Zhang, Z. and Smith, R. L. 2003. On the estimation and application of max-stable processes. Manuscript, Washington University.

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

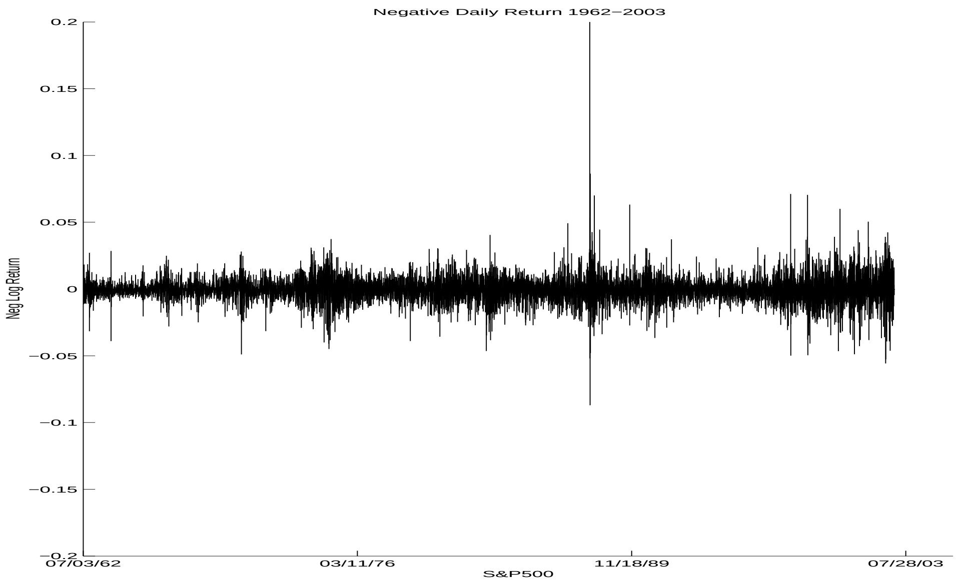

Figure 1: *Plot of the S&P500 original log returns. The highest value corresponds to Oct. 19, 1987 Wall Street crash.*

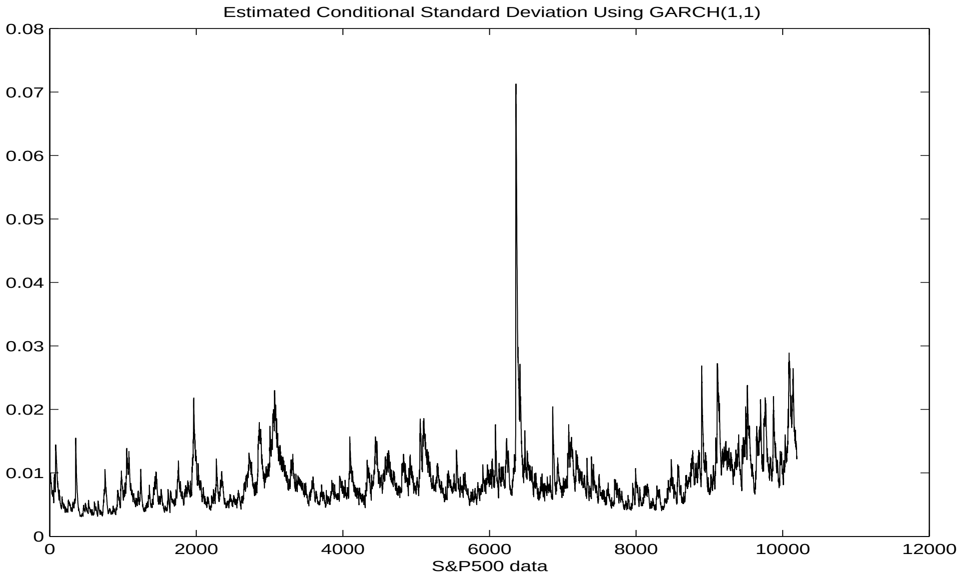

Figure 2: *Estimated volatility plot of the S&P500 log returns. GARCH(1,1) model is used to estimate the volatilities.*

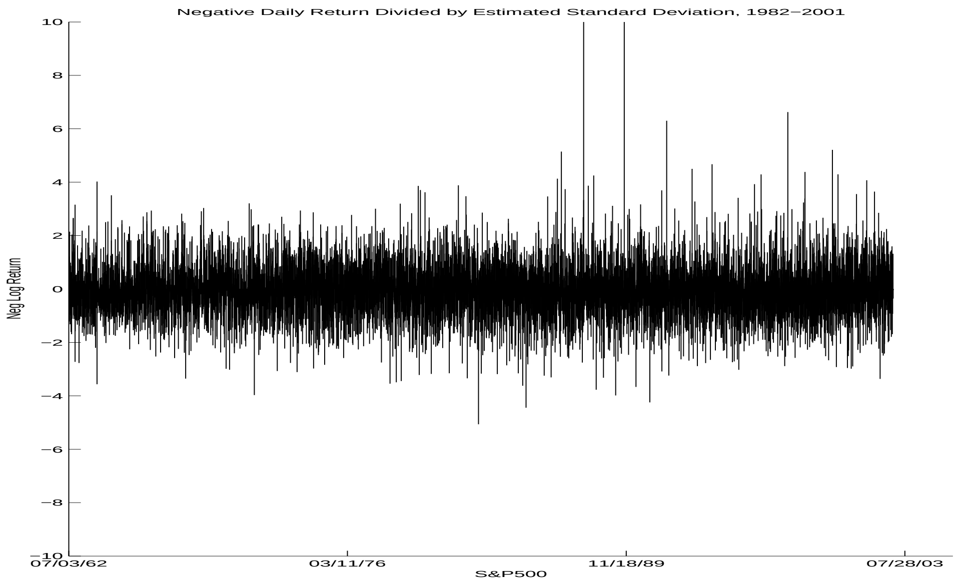

Figure 3: *Plot of the S&P500 standardized log returns.*

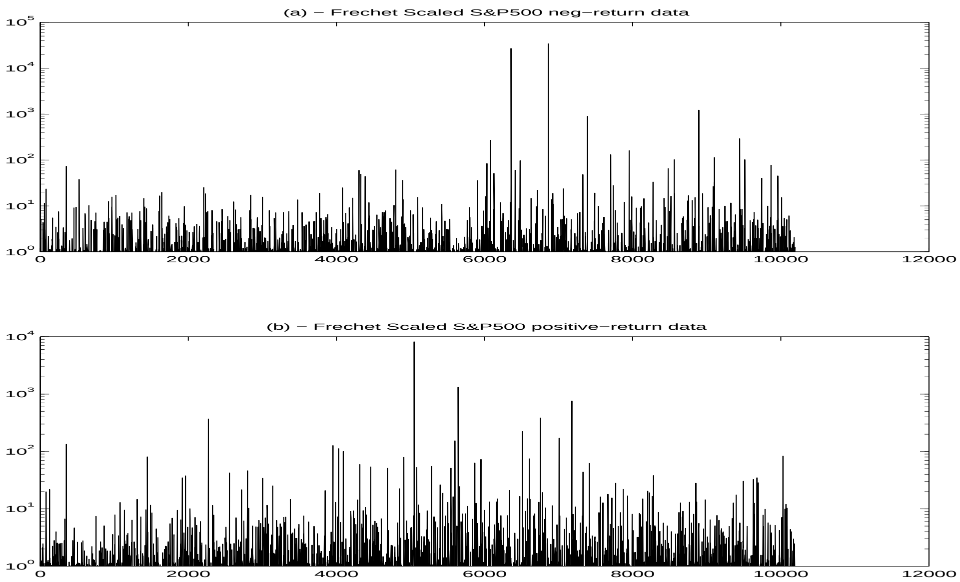

Figure 4: *Plots of the S&P500 standardized time series in Frechet scale. Panel (a) is for negative returns. Panel (b) is for positive returns. The threshold used for both plots is* $u = 1.2$.

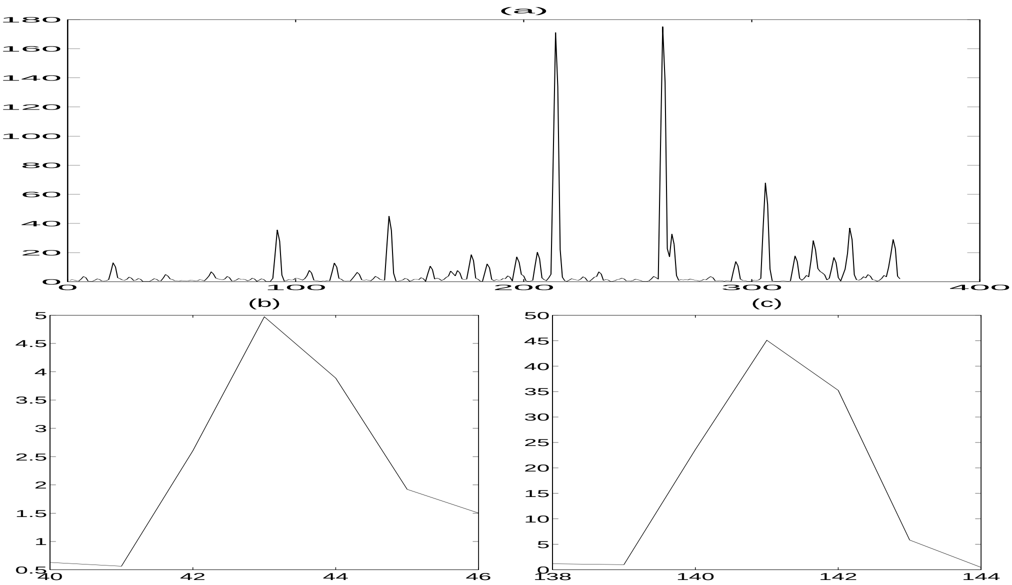

Figure 5: *An illustration of a M4 process. Figure (a) consists of 365 simulated day observation. Figures (b), (c) are partial pictures drawn from (a), i.e. (b) is the enlarged view of (a) from the 40th day to the 45th day; (c) is the enlarged view of (a) from the 138th day to the 144th day. (b) and (c) are showing a single moving pattern, called signature pattern, in certain time periods when extremal events occur.*

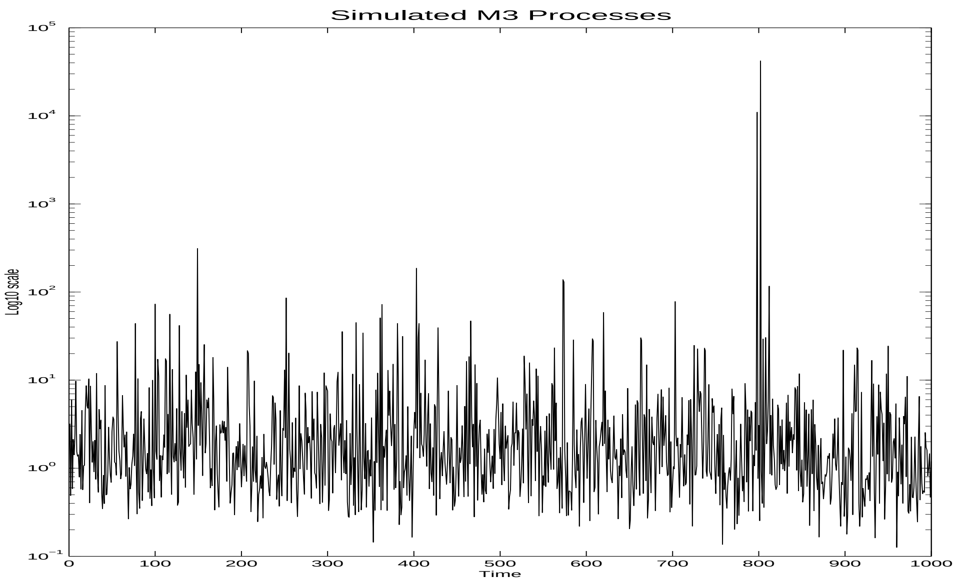

Figure 6: *Simulated time series of M3 processes in model (4.1).*

Series	N_u	μ (SE)	$\log\psi$ (SE)	ξ (SE)
Negative	1050	3.231845	-0.319098	0.096760
		(0.086479)	(0.075502)	(0.028312)
Positive	1088	2.869327	-0.784060	-0.062123
		(0.054016)	(0.064046)	(0.025882)
Absolute	2138	3.514718	-0.450730	0.045448
		(0.071042)	(0.060214)	(0.018304)

Table 1: *Estimations of parameters in GEV using standardized return series and threshold value 1.2. N_u is the number of observations over threshold u.*

lag	Negative returns				Positive returns			
$k=$	Rej.	Ind.	Min	Max	Rej.	Ind.	Min	Max
1	1.0000	0.0700	0.0162	0.3092	0.6787	0.0555	0.0079	0.1291
2	0.7015	0.0700	0.0013	0.2292	0.7556	0.0357	0.0078	0.1905
3	0.6064	0.0525	0.0125	0.2370	0.6454	0.0317	0.0126	0.1630
4	0.8440	0.0525	0.0185	0.4046	0.4104	0.0317	0.0152	0.1170
5	0.5713	0.0306	0.0095	0.3369	1.0000	0.0238	0.0344	0.2113
6	0.4452	0.0088	0.0174	0.2761	0.5147	0.0119	0.0181	0.1597
7	1.0000	0.0131	0.0532	0.2917	0.5125	0.0159	0.0196	0.2245
8	0.7417	0.0306	0.0029	0.1353	0.1158	0.0079	0.0067	0.0485
9	0.3793	0.0175	0.0244	0.3740	0	0.0119	0.0065	0.0268
10	1.0000	0.0131	0.0264	0.0687	0	0.0159	0.0060	0.0186
11	1.0000	0.0088	0.0262	0.2012	1.0000	0.0159	0.0309	0.0444
12	1.0000	0.0131	0.0556	0.1802	1.0000	0.0079	0.0530	0.0679
13	0	0.0175	0.0083	0.0093	0.8681	0.0040	0.0219	0.0320
14	1.0000	0.0088	0.0480	0.0528	0	0.0119	0.0072	0.0081
15	0.5217	0.0131	0.0098	0.0436	1.0000	0.0198	0.0839	0.1061

Table 2: *Columns 2, 6 (×100) yield percentages of rejection of H_0 from all fully enumerated local windows of size 500. Of each data point, at least one component is nonzero. Columns 3, 7 are estimated lag-k tail dependence indexes over a threshold value computed at the 95th percentile for the whole data. Columns 4, 8 are the minima of all computed $\mathrm{Q}_{u,n}$ values using (2.11) in all local windows. Columns 5, 9 are the maxima of all computed $\mathrm{Q}_{u,n}$ values using (2.11).*

r	a_{r0}	SE	r	a_{r0}	SE	r	a_{r0}	SE
0	0.2385	0.2318						
1	0.0700	0.0175	5	0.0306	0.0178	9	0.0175	0.0176
2	0.0700	0.0181	6	0.0088	0.0176	10	0.0131	0.0176
3	0.0525	0.0179	7	0.0131	0.0176	11	0.0088	0.0176
4	0.0525	0.0180	8	0.0306	0.0177	12	0.0131	0.0176

Table 3: *Estimations of parameters in Model (5.1).*

r	a_{r0}	SE	r	a_{r0}	SE
0	0.5913	0.1668	4	0.0314	0.0200
1	0.0550	0.0190	5	0.0236	0.0199
2	0.0354	0.0200	6	0.0118	0.0198
3	0.0314	0.0200	7	0.0157	0.0199

Table 4: *Estimations of parameters in Model (5.2).*