TECHNISCHE UNIVERSITÄT DRESDEN
Fakultät Wirtschaftswissenschaften

Statistics of Multivariate Extremes with Applications in Risk Management

Dissertation
zur Erlangung des akademischen Grades
Dr. rer. pol.

vorgelegt von
Dipl.-Wing. Rodrigo Herrera
Geboren am 6. April 1978
in Talca, Chile

Betreuer: Prof. Dr. Bernhard Schipp
Vorsitzender: Prof. Dr. Udo Broll
Gutachter: Prof. Dr. Hermann Locarek-Junge

Tag der Disputation: 06.06.2009

To Aleksandra,
the greatest gift
is to discover a beautiful heart.

Acknowledges

In the beginning of the present thesis I would like to express my deepest gratitude to Prof. Bernhard Schipp for his excellent guidance and supervision during the entire phase of my doctorate. He filled me with enthusiasm when I more need it. His suggestions and comments have been essential in this thesis. Special thanks to Prof. Dr. Hermann Locarek-Junge for his kind acceptance to be my second referee. I also extend my thanks to my colleagues Enrico Lovaz, Matthias Deutscher, Dr. Sabine Hegewald and Sabine Haller-Schulz for all their help, support, interest and valuable hints.

This research project has been principally supported by a Doctoral Scholarschip President of the Republic of Chile. Further, the Alpha Project grant (SistIng), which has permitted me visiting the Department of Operations Research in the University of Saarland during the year 2007. Especially I would like to thank Prof. Stefan Nickel for giving me the opportunity to spend an extremely fruitful research there. But particularly, I learned that academic work indeed may be lively and funny.

I wish to thank all my friends in Dresden, Chilean and non Chilean, that relationship to this thesis was to stop me thinking and working all the time for it. Though I won't name them all, a big thanks goes to every one of them.

I thank Aleksandra, Agnes, Tom, Anne and Dr. Sabine Hegewald for their efforts and time reviewing my writing many times at the final version of the thesis, correcting and offering suggestions for improvement.

The last lines are for my family, for supporting, motivating and encouraging me during all stages of my life. They are very proud of me, but I am much more of them.

Dresden, Rodrigo Herrera
14th April 2009.

Contents

List of Figures

List of Tables

Nomenclature

$PRM(\mu)$	Poisson random measure with mean μ.
$\mathcal{R}_{-\infty}$	Class of rapidly varying functions.
erfc	Complementary error function.
c.s.m.s	A complete and separable metric space
$\xrightarrow{d}$	Convergence in distribution.
$\xrightarrow{P}$	Convergence in probability.
$\bigvee_{j=1}^{d}$	Denotes the maximum among $j = 1, \dots, d$.
$\max_{1 \le j \le d}$	Denotes the maximum among $j = 1, \dots, d$.
$\bigwedge_{j=1}^{d}$	Denotes the minimum among $j = 1, \dots, d$.
$\min_{1 \le j \le d}$	Denotes the minimum among $j = 1, \dots, d$.
$\mathbb{R}^d$	d-dimensional Euclidean space.
$\mathbb{E}$	Expected value.
EVT	Extreme value theory.
MDA	Maximum domain of attraction.
$\Phi_\alpha(x)$	Frèchet distribution.
GARCH	Generalized autoregressive conditional heteroskedasticity.
GEV	Generalized Extreme Value.
GPD	Generalized Pareto distribution.
$\Lambda(x)$	Gumbel distribution.
$\bigcap_{i=1}^{d}$	Intersection.
$D\left((0, \infty]\right)$	It is the space of real-valued, right continuos functions.
$(A + B)_+$	$\max\left\{(A + B), 0\right\}$.
MEI	Multivariate extremal index.
M_4	Multivariate maxima of moving maxima
$o\,(1)$	$a\,(x) = o\,(b\,(x))$ as $x \to x_0$ means that $\lim_{x \to x_0} a\,(x)\,/b\,(x) = 0$.
POT	Peaks over threshold method.
$\propto$	Proportional to.
$\mathcal{R}_{-\alpha}$	Regular varying with index $-\alpha$.
M_n	Sample maximum.
$2\mathcal{R}$	Second order regular variation.
$\sim$	Similar to.
$M_p\,(\mathbb{C})$	Space of point measures $\mathbb{C}$.
SV	Stochastic volatility models.
$\mathbb{S}^{d-1}$	The (d-1) unit Simplex $\mathbb{S}^{d-1} = \left\{z \in \mathbb{R}^d : \|z\| = 1\right\}$.
$X_{k,n}$	The k-th largest order statistic.

$x \vee y$	The maximum between x and y.
$x \wedge y$	The minimum between x and y.
$F^{\leftarrow}$	The quantile function of the distribution F.
x_F	The right endpoint of a distribution.
$M_+(\mathbb{C})$	The space of Radon measures on a space $\mathbb{C}$.
$\exists$	There exist.
$\bigcup_{i=1}^{d}$	Union.
$\xrightarrow{v}$	Vague Convergence.
$\Rightarrow$	Weak convergence.
$\Psi_\alpha(x)$	Weibull distribution.

CHAPTER 1

Introduction, motivation and new scientific contributions

"Curiouser and curiouser."
(Alice in Wonderland)

1.1. Introduction

The current subprime crisis, together with its consequences for international markets, shows that a deeper understanding of extreme events in statistical data from economics, insurance and finance is of high priority. In a speech in Dublin, Charlie McCreevy, the European Market Commissioner, denounced

> "The irresponsible lending, blind investing, bad liquidity management, excessive stretching of rating agency brands and defective value at risk modelling that prompted the turmoil of recent months" (red. subprime credit crisis). Financial Times, Friday, October 26, 2007.

Comovements and extreme events between financial asset-returns have significantly increased during recent time periods in almost all international markets. Asset prices do not move independently of one another, nor do markets function on a standalone basis.

This thesis presents new results in extreme value statistics with particular emphasis on risk management. Extreme events refer, for example, to extraordinary claims to insurance companies, crashes of equity markets or extreme losses in credit portfolios due to borrower defaults. Hence, extreme events occur rarely, i.e., only few extreme observations are available. Nevertheless, probabilities and dependence structures have to be assigned to extreme events due to their economic impact.

In this dissertation we study a variety of crises and crashes from the perspective of an investor in financial markets. The two key concerns that an investor has regarding crises and crashes are their influence on his risk exposure and their effect on his asset allocation decisions. This thesis analyzes these aspects in several ways.

A crash is defined as an extreme event in a single asset, a single sector or a single market. We consider a crisis as a period with marked extreme dependence or comovements, which affect many assets in an industry, a single market or several markets worldwide. While a crash refers to a specific event in one asset, industry or market on its own, a crisis refers to a period of turmoil in several assets, industries or markets at the same time.

In order to highlight the motivation of this thesis more precisely, we will restrict our attention to the next example. Suppose that the stock market price P_i has been observed, for example the DAX index. The series of the observed price is transformed into log-returns by taking log-differences.

$$X_i = \log P_i - \log P_{i-1}.$$

This transformation is commonly used in finance (see for example Jorion (2003)). There are several reasons for this choice of transformation. The most important one is the belief that log-returns, in contrast to prices, can be understood as realization of a stationary process.

The first question that we would like to answer is

> *What can one say about the extremal behaviour of the distribution of the future losses of this asset?*

Of course, the answer is not easy. There are too few observations to perform a valid statistical analysis. Suppose for the moment that the data are independent and identically distributed (iid). Then, in a classical framework one could use the whole sample to fit a Gaussian or Elliptical distribution and use the tails to determine the future losses. A standard measure of this type is the Value at Risk (VaR): the worst expected loss under normal market conditions over a specific time interval at a given confidence level. For instance, for a given portfolio, a 99% 10-day VaR of one million euro means that the probability of incurring a portfolio loss of one million euro or more by the end of a two-week (10 trading days) period is 1%. However, there exist other more complicated levels to estimate than the 99%. The Basel Committee[1] formulates international capital adequacy for financial institutions, as for example the Economic Capital which is currently at a level of 99.97% per year! It is at this point that extreme value theory (EVT) may enter, offering an alternative formulation based on the asymptotic distribution of the maxima. Recent introductory books on this subject are Embrechts et al. (1997); Coles (2001); Falk et al. (2004); Beirlant et al. (2004).

The normal distribution is the important limiting distribution for sample averages as summarized in a central limit theorem. However, one cannot hope that the magnitude of an extreme event like the crash of 1987 could be modelled by such class of distributions. In fact, under the Gaussian hypothesis for any given stock, an observation more than five standard deviations from the mean should be observed about once every 7,000 years!

Fortunately, the family of extreme value distributions is the one to study the limiting distributions of sample extrema. The Fisher and Tippet theorem suggests that the asymptotic distribution of the maxima of iid random variables under some norming constants $a_n > 0$, and $b_n \in \mathbb{R}$, belongs to the type of one of the following three distribution functions:

$$\begin{aligned}
&\textit{Frechet}: && \Phi_\alpha(x) &&= \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} && \alpha > 0 \\
&\textit{Weibull} && \Psi_\alpha(x) &&= \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases} && \alpha > 0 \\
&\textit{Gumbel}: && \Lambda(x) &&= \exp\{-e^{-x}\}, && x \in \mathbb{R}
\end{aligned}$$

where α is the tail index of the distribution and characterizes the tail behaviour of the distribution function. Alternatively, a more pragmatic approach is to observe that for sufficiently high threshold u, the distribution function of the excess, $X_i > u$, may be approximated by a generalized Pareto distribution (GPD). In fact, as the threshold gets large, the excess

[1] www.bis.org/bcbs/

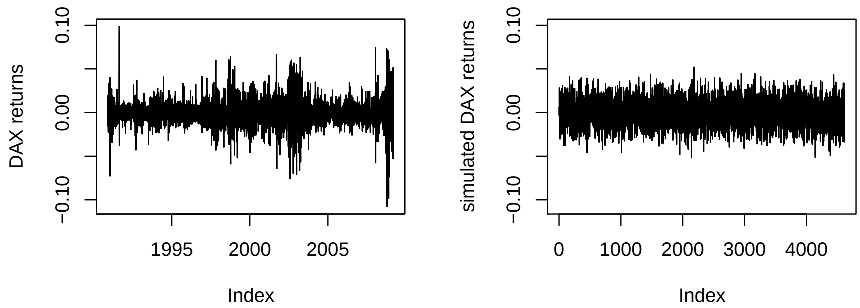

FIGURE 1.1.1. The stock market returns for the DAX index is shown on the left, in the period January 1990 - March 2009. A simulation based on the properties of the DAX returns under the assumption of Gaussian distribution is displayed on the right.

distribution

$$F_u(x) = P\left(X - u \leq x \mid X > u\right),\ x \geq 0,$$

converges to the GPD under the condition that the distribution of the returns belongs to some of the tree types of extreme value distributions, i.e.,

$$\bar{F}_u(x) \approx \bar{G}_{\xi,\beta}(x)$$

where $\bar{G}_{\xi,\beta} = 1 - G_{\xi,\beta}$ with tail index $\xi \in \mathbb{R}^2$ and scale parameter $\beta > 0$. The GPD is defined as

$$\bar{G}_{\xi,\beta}(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right) & \xi \neq 0, \\ \exp(-x/\beta) & \xi = 0. \end{cases}$$

The outcome of that is: To estimate high tail probabilities, we can concentrate more on the limit distribution of the maxima than on the entire sample of a distribution. So far, everything is fine, but the result of more than half a century of empirical studies on financial time series indicates that there exists a set of stylized statistical facts, which are common to a wide set of financial assets.

For instance, the distribution of returns seems to display volatility clustering, which basically means the events tend to cluster in time. This phenomenon has led to the development of the class of GARCH models, pioneered by Engle (1982b), who was awarded a Nobel prize. Figure 1.1.1 on the left depicts this stylized fact for the DAX returns. Notice the enormous difference of the real sequence in comparison to a simulation based on the normal distribution. Thus, a direct application of extreme value theory is not possible.

Fortunately, Leadbetter et al. (1983) have shown that under appropriate mixing conditions, there exists a nondegenerate limiting distribution $\tilde{H}$ for an associated iid sequence $\tilde{M}_n$ of the maxima of the stationary sequence M_n. Then, the normalized maximum $\left(M_n - b_n\right)/a_n$ of the dependent series also has a nondegenerate limit distribution H, and they are related to each other by $H = \tilde{H}^{\theta}$, where θ is the *extremal index*. The extremal index quantifies the strength of dependence between threshold exceedances $X > u$, with $\theta = 1$ corresponding to

[2] In fact it is equivalent to $\xi = 1/\alpha$ in the Fisher and Tippet theorem.

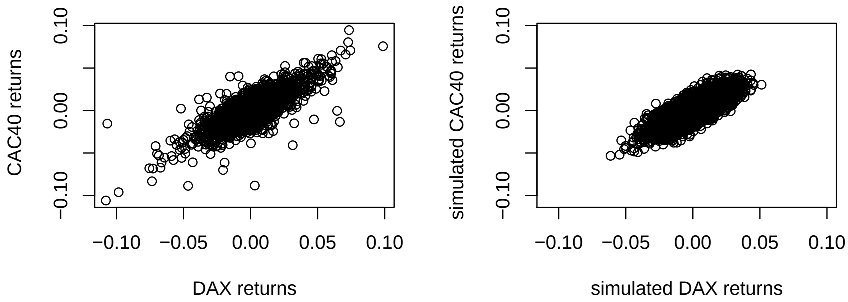

FIGURE 1.1.2. On the left bivariate returns of the DAX and CAC 40 indices. On the right a simulation of these indices under the assumption of that a bivariate normal distribution holds.

asymptotic independence and $\theta \downarrow 0$ to an increasing propensity of large observations to occur in clusters. Assuming that we can calculate this coefficient, we could decluster the returns, i.e., obtain a pseudo iid sample, and apply the classical approach. However, the estimation of this parameter is not an easy task due to different forms of extremal cluster behaviour and the small number of extreme events in the sample (see for instance Laurini and Tawn (2003); Ferro and Segers (2003)).

A common alternative is to utilize parametric models for volatility, as for example, the generalized autoregressive conditional heteroskedasticity family (GARCH) of Engle (1982a) or the stochastic volatility family of Clark (1973). In these models volatility is usually extracted from daily squared returns, which are unbiased but noisy estimates of daily conditional volatility. However, Breidt and Davis (1998); Davis and Mikosch (2006b) demonstrate that there exists no extremal clustering at the extremes for stochastic volatility in both the light and heavy tailed cases. More precisely, the large sample behaviour of maxima is the same as that of the maxima of the associated iid sequence. In the case of GARCH processes, Davis and Mikosch (2006a) have shown that these always exhibit clustering at the extremes. So while both stochastic volatility and GARCH processes exhibit volatility clustering, only GARCH has clustering of extremes. See for instance the contribution of Thomas Mikosch in Finkenstädt and Rootzén (2004) for more detailed information about the behaviour of this class of processes.

For these reasons, it is an advantage to have techniques that are focused purely on extreme movements and are not influenced by the degree of volatility dependence in more routine circumstances. In Chapter 3 we will introduce a model where the data will be fitted in a single step and will not involve the prefiltrate of the volatility to obtain pseudo iid data.

In a multivariate framework the things become even more complicated, because of the fact that the comovements of assets become stronger when markets are under stress. Consider now a bivariate example where the returns are from CAC 40 and DAX.

We are interested in the extreme behaviour in both indices, as for example, it might be of interest to answer questions like:

> *What is the probability of a joint crash in one of the stock markets conditional on the collapse of the other?*

This question is more complicated than the first one, because of the dependence between the two assets. We do not know the type of dependence, so that we should try to infer this from some mode. The workhorse of portfolio theory applications has been the multivariate Gaussian distribution and its variance as the measure of risk. Indeed, standard portfolio selection is usually based on this approach, the Markowitz mean-variance theory and the Sharpe-Lintner-Mossin capital asset pricing model (CAPM). Further, the dependence structure of the asset returns is solely described by the Pearson correlation of the returns. This suffices in case the asset return distributions are multivariate normal. However, if we leave the Gaussian or more generally the elliptical world, a mere consideration of the correlation matrix often explains the dependence structure in a quite unsatisfying way. Pitfalls like a non-existing correlation or zero correlation for dependent random variables may occur. Especially the dependence structure of extreme events is usually poorly or incorrectly described by this measure.

In Figure 1.1.2 we display a scatter plot for the returns of the CAC 40 and DAX indices; on the left we have the true returns, while on the right we have a simulation of these returns under the assumption of a bivariate Gaussian distribution. As we can observe extreme events in the simulated data are scarce. In fact, the multivariate Gaussian distribution suffers of asymptotic independence, which means that the probability of a simultaneous extreme event in two o more marginals is equal to zero. Hence, this is not appropriate to model extreme dependence, for example, among stock markets.

A more appropriate framework is multivariate extreme value theory, which has been developed for studying the joint distribution of extremes in several series. The probabilistic limit theory of multivariate extremes is reasonably well established, and has been reviewed in the books of Galambos (1975); Resnick (1987); Coles and Tawn (1991); Falk et al. (2004), but the statistical theory is still in rapid development. In contrast to the univariate case, the multivariate extreme value distributions cannot be represented by a parametric family because the class of dependence structures is simply too large. Instead, the family of multivariate extreme value distributions can be indexed by a class of finite measures (exponent measures; Starica (1999); Heffernan and Resnick (2005); Resnick (2006); Balkema and Embrechts (2007)) or in another description by a class of dependence functions (Pickands dependence functions or Copulas functions; Kotz and Nadarajah (2002); Klüppelberg and Mayer (2006) or Joe (1997); Embrechts et al. (2003); Nelsen (2006), respectively). Although the applications in the real world are endless in financial econometrics and risk management, these are hindered by the lack of data and rigorous techniques for high dimensions.

The Copula methodology has recently come to the attention in various sciences as a way to overcome the limitations of classical dependence measures as the linear correlation, see for instance Embrechts et al. (2002) for a directory of pitfalls of the correlation as measure of extreme dependence. Standard literature in copulas and their applications are the books of Nelsen (2006); McNeil et al. (2005).

In brief, for multivariate distributions we can separate the marginals from the dependence structure which is represented by the copula function. This leads to a two step modelling

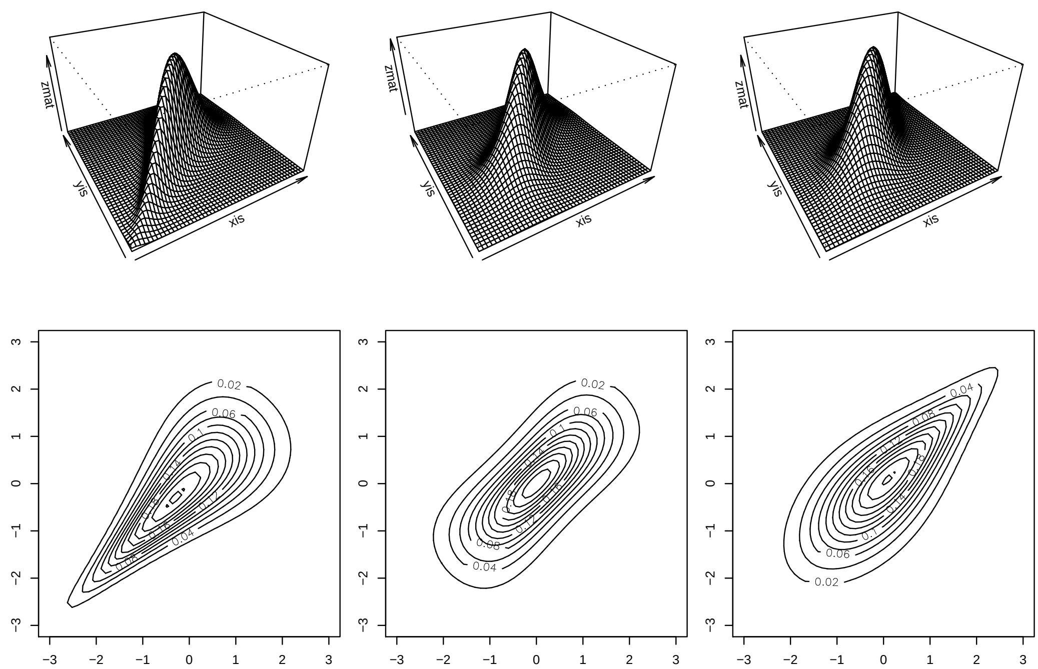

FIGURE 1.1.3. Estimated density function copula for the DAX and CAC 40 returns, under the assumption of the copulas: (left) Clayton, (middle) Frank and (right) Gumbel. In the top panel are displayed perspective plots, while in the bottom panel are shown the contour plots.

of multivariate distributions: specify the marginals $F_1, \ldots, F_d$ and a copula function. The problem is that there exists an infinite class of copulas, within there are more or less natural ones, just like in the case of general multivariate distributions and, like in our case, multivariate extreme value distributions. For the returns of the DAX and CAC 40 indices, for example, we can have the three copulas displayed in Figure 1.1.3, whose linear correlation is the same. This situation and many others have led to some heated discussions between those in favour and those against copula thinking. See for instance Mikosch (2006) followed by a lively discussion and rejoinder in the same volume.

The dependence estimation is a problem but not the only one. Volatility clustering is also present in the multivariate setting. The notion of the extremal index can be extended to a multivariate case, but it is relatively more complicated, because this parameter is now a function. A first definition of this concept was given by Nandagopalan (1994).

In order to apply multivariate extreme value theory some form of data filtering it is necessary to ensure independence of the observations in each marginal. A few declustering schemes have been proposed for multivariate sequences as for example Coles and Tawn (1991) or Nadarajah (2001). However, in each case we need to choose one or more arbitrary parameter, which influence drastically the estimations.

Summarizing, extreme value theory offers a powerful and key theory for use in risk management. Indeed, just in few fields of application we have statistical quantities hard-wired in the law, like VaR–based capital requirements within the Basel II regulatory framework for large international banks. However, the impact of all these developments on day to day risk management has been minimal so far. The case where the data are not iid, the complexity of multivariate extensions and the visualization of extreme events has been highlighted in this introduction. The next section summarizes the main contributions of this thesis to solve these types of problems.

Outline of the thesis and new scientific contributions

The contributions of this thesis have mainly a dual purpose: introducing several multivariate statistical methodologies where in the majority of the cases only stationary of the random variables is assumed and also highlighting some of the applied problems in risk management where extreme value theory may play a role. Mostly every chapter is self-contained, because they have their own more detailed introduction and short conclusion. The following is a brief scheme of the thesis.

Chapter 2: "Univariate and multivariate extremes"

This chapter provides the reader with a basic background on extreme value theory in the iid and stationary case, in the univariate and multivariate framework. Further, it introduces the main differences between the iid and stationary case. We explain some theoretical concepts in extreme dependence, as for example the influence of clustering extremes on the limiting behaviour of the maxima. In addition, we give a short application containing several new results about the extreme dependence among different marginals, when the data are not iid. The main conclusion of this chapter is that the results obtained for different measures of extreme dependence can take very different values, when we work only with the associated ii random vector instead of the true stationary sequence.

Chapter 3: "The Multivariate extremal index and the visualization of measures of extreme dependence"

This chapter concentrates principally on the estimation of the multivariate extremal index, cluster size probabilities and visualization of extreme dependence in the non iid case. First, a new approach is proposed in order to estimate the multivariate extremal index. This approach does not require an arbitrary choice of declustering parameters based on a property of the time of exceedances over a sequence of thresholds. Second, we discuss a new declustering method based on a priori estimation of the multivariate extremal index, so that the associated declustered time series, i.e., a pseudo iid of maxima, are supported by the limiting theory. Third, a novel measure of tail dependence introduced by Hsing et al. (2004) is analyzed and some asymptotic properties are derived. Finally, we use this framework to analyze the financial crisis in South East Asia in late 1997. We find that, prior and posterior to the crisis, there was not a marked extreme dependence among the Asian markets, Japan and U.S.A. This chapter gives a first feeling how the cluster at the extremes behaves in each marginal and shows the relation of this with the extreme dependence among the marginals. However, the methodology states in this chapter tries to find the best associated iid sequence of the stationary time series, which is not a desirable characteristic. In addition, the nonparametric approach does not allow making inference about future extreme events.

Chapter 4: "Multivariate threshold methods with self-exciting functions"

In Chapter 4 we address these limitations by introducing a generalization of the Peaks over threshold model for the stationary case, where the data will be fitted in a single step and the observations are not treated anymore as iid. The main aims of the models are based on marked point processes combined with self-exciting processes, where the intensity of arrival of extreme events can depend on past extreme events, allowing more realistic models. The innovating feature of the present chapter is that, contrary to earlier works on multivariate extreme models, the temporal dependence structure of each underlying marginal is not treated with an ARCH, GARCH, Multifractal Random Walk etc., filter, which certainly impacts on the dependence properties among the marginals. The second contribution in this chapter is the concept of dynamical dependence among marginals, which has been recently introduced in risk management by Dias and Embrechts (2004), Patton (2006) and Giacomini et al. (2007) for instance. In all these papers, the authors investigate the dynamical evolution of the multivariate dependence through the existence of structural changes in the dependence. We follow these authors, however, under the point of view of the Pickands dependence function and not of copulas for the multivariate extreme dependence. We propose two parametric models to take into account the multivariate extreme dependence and a semiparametric model in the spirit of the model Dias and Embrechts (2004), but with other structure to determine the time of change of the multivariate dependence, which has the characteristic of time-invariant. The third important contribution of this chapter is the application of these models to the three most important stock market indices from U.S future markets, the S&P 500, the Dow Jones and the Nasdaq indices. Different examples of stress testing scenarios were calculated to the most important extreme events, as the mini crash of 1997, the Russian default of 1998, the Dot.com crash of 2000, and the extreme movements by the attacks on September 11, 2001. The models proposed in this section are an important contribution to the multivariate extreme value literature. Certainly, as in most works in the multivariate framework, statistical estimations are limited to low dimensions, because of the fact that for d dimensions we need to calculate 2^d derivates of the maximum likelihood function to obtain all combinations of dependence among the marginals.

Chapter 5 : "On the estimation of M_4 processes through mixture Dirichlet process models"

This chapter addresses the dimensionality issue based on a wider context of extreme value theory; the infinite dimensional generalization of extreme value theory, which leads to max-stable processes, introduced by De Haan (1984). These processes have the potential to describe clustering behaviour. One of the most important features of max-stable processes is that they do not only model the dependence among the marginals, but also model the dependence across time. In particular, we concentrate on a class of max-stable processes introduced by Smith and Weissman (1996) to characterize the joint distribution of extremes in multivariate time series. The Multivariate maxima of moving maxima process (M_4 processes for short). The M_4 processes turn out to constitute a rich subclass of those of general multivariate stationary processes, mainly because those processes have the same multivariate extremal indexes as the

M_4 processes, i.e., the behaviour of the cluster of extremes in the multivariate framework is the same.

The major reason to investigate these processes in this thesis is that if we are able to estimate a finite version of the M_4 processes, we will be capable to approximate max-stable processes, i.e., multivariate extreme value processes, whose representation allows us to approximate multivariate time series with clusters at the extremes and heavy tail behaviour. However, the estimation of these classes of processes is a challenging problem in itself, because of the fact that they suffer from degeneracies, i.e., the joint density is typically singular with respect to the Lebesgue measure and this causes problems for maximum likelihood techniques. Smith (2003) showed that these singularities will hold infinitely often if we can observe the process for a long period of time. This process will create a multidimensional deterministic pattern, which yields to the determination of the coefficients of the M_4 process. The problem is then to know how many of these singularities exist.

The contribution of this chapter is a nonparametric Bayesian approximation to the estimation of M_4 processes. The idea is to estimate these singularities in the multivariate density by means of an infinite mixture of Dirichlet processes of Gaussian distributions based on Markov chain Monte Carlo methods. Until now, only approximations based on the experience or arbitrary choices have been used. In this chapter we allow to deduce this number of singularities from the data. The results of the experiments show that the mixture of Dirichlet processes of Gaussian distributions can give an accurate estimation of the signature patterns in a M_4 process. We observe further that the estimation of the exact number of singularities in a M_4 process depends more on the number of exceedances over a threshold $\boldsymbol{u}$, while the number of dimensions plays a second role.

A second contribution of this chapter is the application of M_4 process in the context of risk management, incorporating both clustering and heavy tails to the analysis of some selected German stock markets. In a first approach we illustrate an application to portfolio optimization under VaR constraints. We compare M_4 processes with the two approaches variance-covariance and historical simulations. The VaRs computed from the M_4 approach are higher in general. Moreover, the results indicate that the M_4 approach mimics the cluster behaviour at the extremes of the original sequence. In a second application, we establish different "perfect storm" scenarios to stress test different portfolios obtained in the first application. The scenarios given by the M_4 approach are more realistic in case of the recent crisis. The best hedging is also given by this approach.

Chapter 6: "Topics in multivariate regular variation in finance"

Chapter 6 focuses on two different problems. Recent research in risk management has highlighted the importance of the conditional correlation as measure of contagion or interdependence, and the estimation of spillover probabilities in some financial crisis. This chapter will offer some insight into these issues from the point of view of the theory of multivariate extremes. A first contribution is to demonstrate how the conditional correlation conduces to erroneous conclusions in extreme events. We show through a simple linear factor model, as it is done in the literature of contagion (Baig and Goldfajn (1999); Forbes (2002); Arestis et al. (2005)), that the assumption of a specific distribution function for the common risk factor affects the conditional correlation directly; hence this is not a reliable metric of dependence. In particular, we concentrate on regular varying distributions with index $\alpha > 2$, which is equivalent to distributions whose maximum domain of attraction (MDA) of the maxima is the Frèchet distribution $\Phi_\alpha = \exp\{-x^{-\alpha}\}$ and an extension for distributions whose MDA for the maxima is in the domain of the Gumbel distribution $\Lambda = \exp\{-\exp\{-x\}\}$.

Furthermore, we provide theoretical arguments suggesting that the strong conclusion of “contagion” or “interdependence”, obtained by the literature based on bivariate correlation tests, follows from arbitrary assumptions, especially on the distribution of the random variable for the common risk factor, which biases the results obtained.

Up to this part in the thesis we have concentrated on the use of asymptotically dependent random variables, which have wide use in financial applications. However, if the series are truly asymptotically independent, i.e., the conditional probability of one variable given that the other is extreme is nearly zero, such an approach will result in the over estimation of extreme value dependence, and consequently in the measure of extreme risk.

Therefore, a second contribution is a semi-parametric model estimation based on power transformations of the marginals and a scaling property of exponent measures for the asymptotic dependence and independence case in multivariate extreme value theory for the iid case. The concept of asymptotic independence used in this chapter is stronger than the asymptotic independence used generally in extreme value theory, but weaker than independence. Another contribution of this chapter is to examine the possible contagion during the Russian crisis until the more recent subprime crisis between Brazil and Russia.

Our empirical investigation reveals some degree of asymptotic independence among these markets, which should implicate no significant contagion effect between bond and stock exchange markets of Russia to Brazil. Nevertheless, there exist significant spillover probabilities during the Russian default, and the more recent crisis.

The software tools developed for the various analyses, based on R, Matlab and C, are not presented in the thesis. Interested readers may obtain more details from the autor (Rodrigo.Herrera@tu-dresden.de).

CHAPTER 2

Univariate and multivariate extremes

"Improving the characterization of the distribution of extreme values is of paramount concern."
(A. Greenspan, 1985)

2.1. Introduction

The financial industry, including banking and insurance, is undergoing major changes. An increasing complexity of financial instruments calls for sophisticated risk management tools. Extreme value theory plays an important methodological role within risk management for insurance, reinsurance, and finance. The question one try to answer is: "If things go wrong, how wrong can they go?" The variance used as a risk measure is unable to answer this question. Alternative measures regarding possible values out of the range of available information need to be defined and calculated. Extreme value theory (EVT) provides the tools to model the asymptotic distribution of the maximum of a sequence of random variables X_i, and in this sense this theory can be very helpful in order to get a first impression about how wrong things can go. A deeper insight into EVT allows knowing not only the order of convergence of the maximum but also the limiting distribution of the largest observations of the sequence. These observations are the main ingredients of more informative risk measures that are normally utilized, like Value at Risk (VaR) or Expected Shortfall.

In this section we review the basic background for univariate and multivariate extreme value theory for iid random variables, as well as the corresponding theory for stationary processes. There is a rich literature on extreme value theory that goes back to the 1920s. Recent introductory books for the univariate case on the subject are Coles (2001), which emphasizes statistical modelling, and Embrechts et al. (1997), which is a comprehensive reference for the theory and its applications to insurance and finance. Leadbetter et al. (1983) is mostly concerned with extremes of stationary processes.

The multivariate theory is naturally more recent. Useful representations in terms of max-stable distributions, regular variation functions, or point processes, have been established. For instance, Resnick (1987) is a key reference the point process theory applied to extreme value analysis, while Resnick (2006) gives detailed information about the relation of multivariate regular variation and multivariate extremes. Works on stationary multivariate extremes have been mainly due to Hsing et al. (1988); Hsing (1993); Nandagopalan (1994), generalizing the main results in stationary sequences by Leadbetter et al. (1983, Chapter 3, pp. 51.). For readers who are not familiar with extreme value theory and its extensions, we provide in this chapter a short outline of these concepts.

2.2. Extreme value theory in the univariate case

2.2.1. Extreme value theory for iid sequences. Suppose $X_1, \dots, X_n$ be a sequence of iid non degenerate random variables with distribution function F. The primary concern of extreme value theory relates to extreme values or maximal values as for example, large asset returns or large portfolio losses. Rewriting the sample in the non-increasing order we get

$$X_{1,n} \geq \cdots, \geq X_{n,n} \tag{2.2.1}$$

Define $M_n = X_{1,n}$ as the sample maximum, where (2.2.1) are the sample order statistics and $X_{k,n}$ is the k- th largest order statistic. Then, M_n has the distribution function

$$\mathbb{P}\left(M_n \leq x\right) = F^n(x).$$

We can see that the maximum of a sample tends to the right-hand endpoint of the distribution almost surely, no matter whether it is finite or infinite. Let x_F be the right endpoint, denoted by

$$x_F = \sup\left\{x \in \mathbb{R} : F(x) < 1\right\}.$$

We immediately obtain, for all $x < x_F$,

$$\mathbb{P}\left(M_n \leq x\right) = F^n(x) \to 0, \quad n \to \infty,$$

and, in the case $x_F < \infty$, we have for $x \geq x_F$ that

$$\mathbb{P}\left(M_n \leq x\right) = F^n(x) = 1.$$

Thus, $M_n \xrightarrow{P} x_F$ as $n \to \infty$, where $x_F \leq \infty$. Since the sequence M_n is non decreasing in n, it converges asymptotically, and hence we conclude that

$$M_n \xrightarrow{a.s.} x_F, \quad n \to \infty.$$

However, this is not immediately helpful, since the distribution function F is unknown and it does not guarantee the existence of a non degenerate distribution. An alternative approach is to look for approximate families of models for F^n. This means, that we are interested in limits of the form

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} \mathbb{P}\left(a_n^{-1}(M_n - b_n) \leq x\right) = H(x) \tag{2.2.2}$$

for suitable normalizing constants $a_n > 0$, and b_n. First of all it is essential to know under what conditions of F there is a nontrivial limit of $\mathbb{P}\left(M_n \leq x\right)$ as $n \to \infty$ for appropriate sequences of constants u_n.

THEOREM 2.2.1. *(Poisson approximation). Let $\tau \in [0, \infty]$ and a sequence $u_n = u_n(x) = xa_n^{-1} + b_n$ of real numbers the following are equivalent.*

$$n\overline{F}(u_n) \to \tau, \tag{2.2.3}$$

$$\mathbb{P}(M_n \leq u_n) \to e^{-\tau}, \tag{2.2.4}$$

where $\overline{F} = 1 - F$ is the tail of F.

We can see if we take logarithms in (2.2.4) we get

$$-n\ln(1-\overline{F}(u_n)) \to \tau,$$

since $-\ln(1-x) \sim x$ for $x \to 0$, this implies that $n\overline{F}(u_n) = \tau + o(1)$, giving (2.2.3). In the second case

$$\mathbb{P}(M_n \leq u_n) = F(u_n)^n = \left(1 - \overline{F}(u_n)\right)^n = \left(1 - \frac{n\overline{F}(u_n)}{n}\right)^n,$$

which implies (2.2.4). Conversely, if (2.2.4) holds for some τ, then (2.2.3) holds. It is also important to know whether there can be sequences of constants $a_n, \tilde{a}_n > 0$ and $b_n, \tilde{b}_n \in \mathbb{R}$ such that $a_n^{-1}(M_n - b_n)$ and $\tilde{a}_n^{-1}(M_n - b_n)$ converge in distribution to two different random variables with very different distributions H and $\tilde{H}$. Fortunately, the following result states that any two possible such distribution functions H and $\tilde{H}$ are closely linked.

PROPOSITION 2.2.2. *(Convergence to types Theorem). Let $\xi, \tilde{\xi}, Y_n$, $n \geq 1$, be random variables such that neither ξnor $\tilde{\xi}$ are almost surely constant, and let $a_n, \tilde{a}_n > 0$ and $b_n, \tilde{b}_n \in \mathbb{R}$ be constants.*

(1) *If*

$$a_n^{-1}(Y_n - b_n) \xrightarrow{d} \xi \quad \text{and} \quad \tilde{a}_n^{-1}(Y_n - \tilde{b}_n) \xrightarrow{d} \tilde{\xi}, \tag{2.2.5}$$

then exist $A > 0$ and $B \in \mathbb{R}$ such that

$$\tilde{a}_n/a_n \to A \quad \text{and} \quad (\tilde{b}_n - b_n)/a_n \to B \tag{2.2.6}$$

and such that

$$\tilde{\xi} \overset{d}{=} (\xi - B)/A. \tag{2.2.7}$$

(2) *If (2.2.6) holds, then either of the two relation in (2.2.5) implies the other and (2.2.7) holds.*

PROOF. See Resnick (1987), Proposition 0.2, pp. 7. □

If (2.2.2) holds, we say F belongs to the maximum domain of attraction of H and we written $F \in MDA(H)$. The following result is the corn stone of classical extreme value theory.

THEOREM 2.2.3. *(Fisher-Tippett Theorem, limit laws for maxima). Let X be a sequence of iid random variables. If there exist norming constants $a_n > 0$, and $b_n \in \mathbb{R}$ and some non-degenerate distribution function H such that*

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} H, \tag{2.2.8}$$

then H belongs to the type of one of the following three distribution functions:

$$\text{Frechet}: \quad \Phi_\alpha(x) \quad = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \alpha > 0$$

$$\text{Weibull} \quad \Psi_\alpha(x) \quad = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0$$

$$\text{Gumbel}: \quad \Lambda(x) \quad = \exp\{-e^{-x}\}, \qquad x \in \mathbb{R}.$$

PROOF. see Resnick (1987), Proposition 0.3, pp. 9. □

The Frèchet, Weibull, and Gumbel distributions are also know as the *Extreme Value Distributions.* This class of distributions provide us with techniques to trade off the bias of having insufficient data in practice and meaningful extrapolations beyond the range of given data. Frèchet distributions are primarily applied in finance because of their unbounded support on the positive halfline and their relationship to so called heavy-tailed distributions. Note that the tail index α characterise the extreme value distribution Frèchet and Gumbel. This index is also know as the index of regular variation.

DEFINITION 2.2.4. A distribution tail $\overline{F}$ is regularly varying with index $-\alpha$ for some $\alpha \geq 0$, we write $\overline{F} \in \mathcal{R}_{-\alpha}$, if

$$\lim_{x\to\infty} \frac{\overline{F}(xt)}{\overline{F}(x)} = t^{-\alpha}, \ t > 0.$$

A complete definition of regularly varying functions and their properties most important can be found in Bingham et al. (1987). A more precise definition together with some application will be given in Chapter 6.

A possible characterization, which combine the three extreme value distributions into a single distribution, is known as the *Generalized Extreme Value* (GEV) distribution and is defined as

$$H_{\xi,\psi,\mu}(x) = exp\left\{-\left(1+\xi\frac{x-\mu}{\psi}\right)_+^{\xi^{-1}}\right\}, \tag{2.2.9}$$

with parameters $\xi \in \mathbb{R}$,$\psi > 0$,$\mu \in \mathbb{R}$, and where $z_+ = \max(z, 0)$.

The parameter ψ and μ are scale parameters and location, respectively. The shape parameter ξ determines the tail behaviour of the GEV distribution. GEV distributions with $\xi > 0$ correspond to the Frèchet distribution with $\alpha = \xi^{-1}$, the limiting case $\xi \to 0$ corresponds to Gumbel distribution. The last case $\xi < 0$ corresponds to the Weibull distribution with $\alpha = -\xi^{-1}$. An intuitive interpretation is that $\xi > 0$ corresponds to a long-tailed distribution; the limit $\xi \to 0$ to a distribution with exponential-type tail; and $\xi < 0$ to a short-tailed distribution with finite upper endpoint.

2.2.2. Domains of attraction. We say that a random variable X with distribution function F belongs to the maximum domain of attraction of an extreme value distribution H if relation (2.2.8) holds. In that case we write $F \in MDA(H)$. We get the following characterizations of the maximum domain of attraction of an extreme value distribution H.

THEOREM 2.2.5. *(Characterization of $MDA\left(\Phi_{\alpha}(x)\right)$)*
The distribution function F belongs to the maximum domain of attraction of $\Phi_{\alpha}(x)$ $(\alpha > 0)$ if and only if $\overline{F}\,(x) \in \mathcal{R}_{-\alpha}$. In that case

$$a_n^{-1} M_n \xrightarrow{d} \Phi_{\alpha}$$

with $a_n = F^{\leftarrow}\,(1 - 1/n)$, where $F^{\leftarrow}$is the quantile function.

PROOF. See for instance Resnick (1987), Proposition 1.11, pp. 54. □

By Taylor expansion, we can see that $1 - \Phi_{\alpha}(x) = 1 - \exp\left(-x^{-\alpha}\right) \sim x^{-\alpha}$, as $x \to \infty$, i.e., the tail of decreases like a power law. At this point is where the concept of regular variation play an important role. Indeed, regular variation tells us how far we can move from exact power law behaviour and still remain in $MDA\left(\Phi_{\alpha}(x)\right)$.

THEOREM 2.2.6. *(Characterization of $MDA\left(\Psi_{\alpha}(x)\right)$)*
The distribution function F belongs to the maximum domain of attraction of $\Psi_{\alpha}(x)$) $(\alpha > 0)$ if and only if $x_F < \infty$ and $\overline{F}\,(x_F - 1/x) \in \mathcal{R}_{-\alpha}$. In that case

$$a_n^{-1}\left(M_n - x_F\right) \xrightarrow{d} \Psi_{\alpha}$$

with $a_n = x_F - F^{\leftarrow}\,(1 - 1/n)$

PROOF. See for instance Resnick (1987), Proposition 1.13, pp. 59. □

Noting that $\Psi_{\alpha}\left(-x^{-1}\right) = \Phi_{\alpha}\left(x\right)$ for $x > 0$, we are not surprised that $MDA\left(\Psi_{\alpha}(x)\right)$ and $MDA\left(\Phi_{\alpha}(x)\right)$ are closely related.

THEOREM 2.2.7. *Every max-stable distribution is of extreme value type, i.e. a non degenerate distribution H is max stable, if $H^m(a_n x + b_n) = H(x)$ holds for some constants $a_n > 0$ and b_n for each $m \geq 2$; conversely, each distribution of extreme value type is max-stable.*

PROOF. See for instance Embrechts et al. (1997), Theorem 3.2.2, pp. 121. □

THEOREM 2.2.8. *(Characterization of $MDA\left(\Lambda(x)\right)$)*
The distribution function F with $x_F \leq \infty$ belongs to the maximum domain of attraction of $\Lambda(x)$ if and only if there exists some $z < x_F$ such that F has representation

$$\overline{F}\,(x) = c\,(x) \exp\left\{-\int_z^x \frac{g\,(t)}{a\,(t)} dt\right\},\ z < x < x_F,$$

where c and g are measurable functions satisfying $c\,(x) \to c > 0$, $g\,(x) \to 1$ as $x \uparrow x_F$, and $a\,(x)$ is a positive, absolutely continuous function with respect to the Lebesgue measure) with density $a'\,(x)$ having $\lim_{x \uparrow x_F} a'\,(x) = 0$.

Other characterization possible is there exist some positive function k such that

$$\lim_{x \uparrow x_F} \frac{\overline{F}\,(x + tk\,(x))}{\overline{F}\,(x)} = \exp\left(-t\right),\ t \in \mathbb{R},$$

holds.

PROOF. See for instance Resnick (1987), Proposition 1.4, pp. 43. □

Now we give some examples of MDA.

Example 2.2.9. (Maxima of Cauchy random variables). Let X be a sequence of iid standard Cauchy random variables, i.e. the density function is given by $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$. Using l'Hospitals rule we obtain

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{(\pi x)^{-1}} = \lim_{x\to\infty} \frac{f(x)}{\pi^{-1}x^{-2}} = \lim_{x\to\infty} \frac{\pi x^2}{\pi(1+x^2)} = 1.$$

Then, $\overline{F}(x) \sim (\pi x)^{-1}$ as $x \to \infty$. Hence, for $x > 0$, $\overline{F}(nx/\pi) \sim (nx)^{-1}$ as $n \to \infty$ and

$$\begin{aligned} \mathbb{P}(M_n \le nx/\pi) &= \left(1 - \overline{F}(nx/\pi)\right)^n \\ &= \left(1 - \frac{1}{n}\left(\frac{1}{x} + o(1)\right)\right)^n \\ &\to \exp\left\{-x^{-1}\right\} = \Phi_1(x), \end{aligned}$$

for $x > 0$. Hence, F belongs to the maximum domain of attraction of the Frèchet distribution. The normalizing constants can be chosen to be $a_n = n$ and $b_n = 0$

Example 2.2.10. (Maxima of exponential random variables). Let X be a sequence of iid standard exponential random variables, i.e., the distribution function F of X is given by $F(x) = 1 - e^{-x}$ for $x \ge 0$. Then,

$$\begin{aligned} \mathbb{P}(M_n - \ln n \le x) &= F(x + \ln n)^n \\ &= (1 - \exp(-x - \ln n))^n \\ &= (1 - n^{-1}e^{-x})^n \\ &\to \exp(-e^x) = \Lambda(x), \quad x \in \mathbb{R}. \end{aligned}$$

Hence, F belongs to the maximum domain of attraction of the Gumbel distribution. The normalizing constants can be chosen to be $a_n = 1$ and $b_n = \ln n$.

In practice, the estimation $\xi \in \mathbb{R}$,$\psi > 0$,$\mu \in \mathbb{R}$ in (2.2.9) requires one to decompose the sample into blocks and take the blockwise maxima. A drawback with this approach is the loss of data. A modern theory in extreme value models is to consider the asymptotic distribution of exceedances over a high threshold u, the peaks over threshold (POT) method. In particular this approach offers many advantages from a statistical point of view. The main aim is to fit a generalized Pareto distribution (GPD) to excesses over a high threshold of a random variable, under the condition that sufficient data are available above the chosen threshold.

Suppose $X_1, \ldots, X_n$ are iid with distribution function $F \in MDA(H_\xi)$ for some $\xi \in \mathbb{R}$, where H_ξ is a non-degenerate limiting distribution and $\lim_{n\to\infty} \overline{F}(c_n x + d_n) = -\ln H_\xi(x)$ holds for normalizing sequences c_n and d_n. The excess distribution function of X is given by

$$F_u(x) = P\left(X - u \le x \mid X > u\right), \; x \ge 0.$$

This relation can be rewrite as

$$\bar{F}(u + x) = \bar{F}(u)\bar{F_u}(x).$$

This result conduces to next definition

THEOREM 2.2.11. *(The Generalized Pareto Distribution (GPD))*
Suppose $X_1, \ldots, X_n$ are iid with distribution function $F \in MDA(H_\xi)$ for some $\xi \in \mathbb{R}$. The excess distribution function of X is given by

$$F_u(x) = P\left(X - u \leq x \mid X > u\right), \; x \geq 0.$$

Now by definition a GPD $G_{\xi,\beta}$ with parameters $\xi \in \mathbb{R}$ and $\beta > 0$ has distribution tail

$$\bar{G}_{\xi,\beta}(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right) & if\ \xi \neq 0, \\ \exp(-x/\beta) & if\ \xi = 0, \end{cases}$$

and $x \in D(\xi, \beta)$

$$D(\xi,\beta) = \begin{cases} [0, \infty) & if\ \xi \geq 0, \\ [0, -\beta/\xi] & if\ \xi < 0. \end{cases}$$

As $F \in MDA(H_\xi)$ then for an appropriate positive function β

$$\lim_{u \uparrow x_F} \sup_{0<x<x_F-u} \left|\bar{F}(u) - \bar{G}_{\xi,\beta}(x)\right| = 0,$$

where x_F is the right endpoint.

This limit is known as Pickands-Balkema-de Haan Theorem. Thus, for high threshold u one expects that the excess distribution F_u can be well approximated by a GPD.

$$\bar{F}_u(x) \approx \bar{G}_{\xi,\beta}(x).$$

The Peaks Over Threshold method can be represented as a semi-parametric model. The excesses above a high threshold u are distributed according to a GPD, while the empirical distribution function of F, or any other appropriate model, is used under the threshold u. This is the semi-parametric extremal model, see for example Coles and Tawn (1991).

This method of analyzing the extreme values has the advantage of being straightforward to implement, but there are a number of disadvantages when considering broader features of the distributions, such as non stationary effects, trends and seasonality into the model. A more substantial variant, introduced by Smith (1989), is via a point process representation of the exceedances. The idea is to view all events exceeding a given level u as a bidimensional point process. We will address this representation in Chapter 4.

2.2.3. Extreme value theory for the stationary sequences. One of the natural generalizations is of an iid sequence to stationary process. We concentrate here on stationary sequences where the dependence is restricted by different distributional mixing conditions. We give conditions on the stationary sequence X which ensure that is sample maxima M_n and the corresponding maxima $\widetilde{M}_n$ of an iid sequence $\widetilde{X}$ (*an iid sequence associated with* X), with common distribution function $F(x) = \mathbb{P}(\widetilde{X} \leq x)$ exhibits a similar limit behaviour.

In the following section, we give a brief survey of definitions of mixing conditions and typical weak-dependence conditions concerning mainly with extremes in sequence of dependent random variables.

CONDITION 2.2.12. $D(u_n)$: For any integers p, q and n
$1 \leq i_1 < \cdots < i_p < j_1 < \cdots < j_q \leq n$ such that $j_1 - i_p \geq l$ we have

$$\left| \mathbb{P}\left(\max_{i \in A_1 \cup A_2} X_i \leq u_n \right) - \mathbb{P}\left(\max_{i \in A_1} X_i \leq u_n \right) \mathbb{P}\left(\max_{i \in A_2} X_i \leq u_n \right) \right| \leq \alpha(l, u_n),$$

where $A_1 = \{i_1, \ldots, i_p\}$,$A_2 = \{j_1, \ldots, j_p\}$ and

$$\lim_{l \to \infty} \limsup_{n \to \infty} \sup \alpha(l, u_n) = 0,$$

for some sequence $l = l_n = o(n)$ and u_n is a sequence of thresholds.

Condition $D(u_n)$ ensures that any two extreme events can become approximately independent as n increases when the index sets A_1 and A_2 are separated by a relatively short interval of length $l_n = o(n)$.

With the restriction $D(u_n)$ Leadbetter et al. (1983, pp. 57.) showed that the extremal types Theorem (2.2.3) holds for stationary sequence.

THEOREM 2.2.13. *Let X be a stationary sequence, let M_n, $a_n > 0$ and $b_n \in \mathbb{R}$ be such that $\mathbb{P}\left(a_n^{-1}(M_n - b_n) \leq x\right)$ converges in distribution to a non degenerate distribution function $G(x)$. Suppose that X_n satisfies $D(a_n x + b_n)$ for all real x such that $G(x) > 0$. Then, $G(x)$ is a generalized extreme value distribution.*

Theorem 2.2.13 shows that the possible limit distribution for block maxima from stationary processes are therefore the same as in the independent case, and can be used as models in the same way.

The results given so far have been concerned with the possible forms of the limiting extreme value distributions. We now turn to the existence of such a limit in that we formulate conditions under which for a sequence of thresholds u_n, $n(1 - F(u_n)) \to \tau$ as $n \to \infty$ are equivalent for stationary sequences.

In a dependence restricted to the condition $D(u_n)$ there can be stated the following sufficient result:

THEOREM 2.2.14. *Suppose $u_n(\tau)$ is defined for $\tau > 0$, such that $n(1 - F(u_n)) \to \tau$ and $D(u_n(\tau))$ holds for each such τ. If $\Pr(M_n \leq u_n(\tau*))$ converges for some $\tau* > 0$ then converges for all $\tau > 0$ and*

$$\lim_{n \to \infty} \mathbb{P}\left(M_n \leq u_n(\tau^*)\right) \to e^{-\theta\tau}.$$

where θ is the extremal index.

The extremal index gives a measure of the short range dependence exhibited by the extremes of a process and, in particular, indicates the tendency of the extremes to occur in clusters. This takes values between 0 and 1, where 1 indicates that there are not cluster on the extremes, while a value nearly to zero indicates larger cluster size.

The extremal index has an impact on all aspects of extreme values of the stationary sequence. Leadbetter et al. (1983) show how it influences the asymptotic distribution of normalized partial maxima. In particular they show that θ^{-1} is the limiting mean size of clusters. Hsing et al. (1988) show that the process of the normalized times of exceedances over a high threshold u converges to a compound Poisson process with the mean cluster size being θ^{-1}. Ferro and Segers (2003) show that the interarrivals, in terms of normalized times, of consecutive exceedances of over a high threshold u converges to a random variable which is 0 with probability θ, otherwise, it is exponential with mean θ^{-1}.

DEFINITION 2.2.15. We say that the process X has extremal index $\theta \in [0,1]$ if, for each $\tau > 0$,

(1) There exists $u_n(\tau)$ such that $n(1-F(u_n)) \to \tau$,
(2) $\mathbb{P}(M_n \le u_n(\tau*)) \to e^{-\theta\tau}$.

We conclude immediately that iid sequences whose maximum converges to a non degenerate limiting distribution have unit extremal index. Leadbetter et al. (1983) show that if (2.2.3) holds then condition $D(u_n)$ is also sufficient to guarantee that $\liminf \mathbb{P}(M_n \le u_n) \ge e^{-\tau}$ but a further assumption is needed to obtain the opposite inequality for upper limit. So, they introduced another dependence condition, $D'(u_n)$, defined in the following way.

CONDITION 2.2.16. $D'(u_n)$ is an anticlustering condition and holds for a stationary sequence X if

$$\lim_{k\to\infty} \limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} \mathbb{P}(X_1 > u_n, X_j > u_n) \to 0, \quad \text{as } k \to \infty,$$

then,

$$\mathbb{P}(M_n \le u_n) - \mathbb{P}^n(X > u_n) \to 0 \qquad n \to \infty. \tag{2.2.10}$$

The natural interpretation $D'(u_n)$ is that, joint exceedances of u_n by pairs (X_i, X_j) become very unlikely for large n.

Recall $\tilde{M}_n$ can have a non degenerate limiting distribution only if

$$\lim_{x\to x_F} \mathbb{P}(X \ge x) / \mathbb{P}(X > x) = 1. \tag{2.2.11}$$

In the absence of this condition, the dependence structure is such that large value has a greater chance of being followed by another one. If the time between two consecutive such values is small relative to n, the passage to the limit will merge those two extremes onto the same time. Thus, the limit process is then not a Poisson process but a compound Poisson process: any occurrence can be multiple rather than single. The multiplicity is usually random and is called the *cluster size distribution* $\pi(\cdot)$. We leave its formal definition for later.

Equation (2.2.10) shows that the distribution of M_n can be well approximated even if (2.2.11) does not hold. A stricter condition than $D(u_n)$ is the condition $\triangle(u_n)$ by Hsing et al. (1988).

CONDITION 2.2.17. (Condition $\triangle(u_n)$) Let $\bar{t} = (t_1, \ldots, t_p)$, $u_n(\bar{t}) = \{u_n(t_1), \ldots, u_n(t_p)\}$, and let $\mathcal{F}_{l,q}(\bar{t})$ be the σ-field generated by the events $\{X_j > u_n(t_i)\}$,$l \le j \le q$,$1 \le i \le p$. We define

$$\sup |\Pr(AB) - P(A)P(B)| = \alpha_{n,l}(u_n(\bar{t})),$$

where the supremum is taken over $A \in \mathcal{F}_{1,p}(\bar{t})$,$B \in \mathcal{F}_{p+l,n}(\bar{t})$, $q \ge 1$such that $\Pr(A) > 0$ and $\alpha(l, u_n(\bar{t})) \to 0$ as $n \to \infty$ for some $l_n = o(n)$.

We can see that under the condition $\triangle(u_n)$ the events required to become independent are more numerous and at many levels simultaneously. Condition $\triangle(u_n)$ holds if $\triangle(u_n(\bar{t}))$ is in force for every choice of $o < t_1 < \ldots < t_p < \infty$, $k \in \mathbb{N}$. If condition $\triangle(u_n(\bar{t}))$ holds, then there exist sequences l_n and r_n such that

$$1 \ll l_n \ll r_n \ll n, \; nr_n^{-1}\alpha_n^{2/3} \to 0.$$

This condition is needed to ensure convergence for a stationary two dimensional point process.

2.2.4. Point processes in extreme value theory. The concept of exceedances over thresholds can be embedded into the theory of point processes in a natural way. Point processes techniques are by now an unavoidable tool in modern extreme value theory and its results give a deep insight into the structure and occurrence of extreme. Indeed, this will play an important role in Chapter 4. A detailed introduction to the theory of point processes would be beyond the aim of this thesis. For a concise introduction we therefore refer to Embrechts et al. (1997) and the references therein. For a more extensive survey we refer to Daley and Vere-Jones (2003).

One can simply think of a point process N as a random distribution of points X in a state space E, where E is equipped with the σ-algebra $\mathcal{E}$. For a set A and a given configuration (X_i), the point process N counts the number of $X \in A$.

Definition 2.2.18. (Random measure and point process). Let $M_p(E)$ be the space of all point measures on E equipped with an appropriate σ-algebra $\mathcal{M}_p(E)$. A point process on a space E is a measurable mapping $N \,:\, (\Omega, \mathcal{A}, \mathbb{P}) \to (M_p(E), \mathcal{M}_p(E))$.

One of the most important processes in extreme value theory is the following point process.

Definition 2.2.19. (Point process of exceedances) Let u_n be a real number and let X_i be a sequence of random variables. Then, the point process of exceedances

$$N_n(\cdot) = \sum_{i=1}^{n} \varepsilon_{n^{-1}i}(\cdot)\mathbb{I}\left\{X_i > u_n\right\}, \; n = 1, 2....$$

with state space $E = (0, 1]$ counts the number of exceedances of a threshold u_n.

If the sequence X is supposed to be iid or strictly stationary satisfying the assumptions $D(u_n)$ and $D'(u_n)$ we can show the weak convergence of the sequence N_n of such point processes to a homogeneous Poisson process N on the state space $E = (0, 1]$. This same process can be written in the perhaps more intuitively form as a bidimensional point process, one dimension for the time of the exceedances and other for the size of the exceedances. This approach will be address in Chapter 4.

The most important point processes are those for which $N(A)$ is Poisson distributed. Imagine that we have a point process which is binomially distributed: $B_n = \sum_{i=1}^{n} \mathbb{I}\left\{X_i \in A_n\right\}$ for iid X_i counts number of successes among $X_1, \ldots, X_n$ and $p_n = \mathbb{P}\left(X_1 \in A_n\right)$ is the sucess probability. Then, Poisson's theorem tells us that $B_n \xrightarrow{d} \mathit{Poisson}\left(\lambda\right)$ provided $p_n \sim \lambda/n$. This result suggests the following definition of *Poisson random measure*.

Definition 2.2.20. (Poisson random measure PRM) A point process N is called PRM with mean μ (we write PRM(μ)) if the following two conditions are satisfied:

(1) For $A \in \mathcal{S}$,

$$\mathbb{P}\left(N\left(A\right)=k\right)=\begin{cases} e^{-\mu(A)}\frac{(\mu(A))^{k}}{k!} & if \quad \mu\left(A\right)<\infty,\ k\geq 0. \\ 0 & if \quad \mu\left(A\right)=\infty, \end{cases}$$

(2) For any $m \geq 1$, if $A_1, \ldots, A_m$ are mutually disjoint sets in $\mathcal{E}$ then $N\left(A_1\right), \ldots, N\left(A_m\right)$ are independent random variables.

A basic tool for dealing with the asymptotic theory of extreme values is weak convergence of point processes. Basically, the notion of weak convergence permits to a point process to converge in ways that would otherwise be fundamentally impossible.

THEOREM 2.2.21. *(Weak convergence of a point processes of exceedances, iid case or weakly dependent) Assume that X is a sequence of iid random variables with common distribution function F, and let u_n be thresholds values such that $D(u_n)$ and $D'(u_n)$ hold. Then, we can observe that*

$$n\bar{F}(u_n) = \mathbb{E}(N_n) \to \tau$$

holds for some $\tau \in (0, \infty)$. Then, the point processes of exceedances N_n converge weakly in $M_p(E)$ to a homogeneous Poisson process N on $E = (0, 1]$ with intensity τ, i.e., N is a Poisson random measure with mean measure τ.

PROOF. See Leadbetter et al. (1983), Theorem 3.4.1, pp. 59. □

Dependence can cause clustering of extremes, and the Poisson approximation for $N_n(u)$ may no longer be valid under the conditions $D(u_n)$ and $D'(u_n)$. In contrast to the last theorem we want to show that the limiting distribution of $N_n(u)$ under dependence is necessarily compound Poisson under a mixing condition which is stronger than dependence condition $D(u_n)$. A common approach is to suppose that there exists the limit

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n) = e^{-\tau}. \tag{2.2.12}$$

In the next Theorem we assume (2.2.12), the mixing condition $\triangle(u_n)$ and the condition $D'(u_n)$. It is important to recall that $\triangle(u_n)$ implies condition $D(u_n)$, and condition $D'(u_n)$ follows from equation (2.2.12) if the sequence X is φ-mixing (see Rosenblatt (1971)).

THEOREM 2.2.22. *Let N_n be the exceedances point process corresponding to the level u_n in the stationary sequence X. Suppose that $\triangle(u_n)$ holds for X, assume (2.2.12) for some $\tau > 0$, and that $N_n \xrightarrow{d} N$, for some point process N. The, it is necessarily a compound Poisson point process.*

$$N_n(u_n) = \sum_{i=1}^{\infty} \xi_i \varepsilon \psi_i(u_n) = \sum_{i=1}^{\pi(\tau)} \xi_i$$

where $\{\psi_i\}$ are the points of a homogeneous Poisson process on $[0, \infty)$ with intensity $\theta\tau$ and $\{\xi_i\}$ are the iid cluster sizes of N with distribution $\{\pi_i\}$ on $\mathbb{N}$.

PROOF. See Hsing et al. (1988). □

The random variable $\pi\left(\cdot\right)$ is called the limiting cluster sizes if (2.2.12) holds and

$$\pi_n(j) = \mathbb{P}\left\{N_r(u_n) = j \,|N_r(u_n) > 0\right\},$$

for some integer $j \geq 1$ and some sequence $r = r_n$ of natural numbers such that $n \gg r_n \gg s_n \gg 1$. Simply speaking, π_n is the rate of arrival of cluster, where a cluster is defined as the set of exceedances of the threshold u_n that occur in an arbitrary block of length r_n, given that at least one exceedance occurs in the block. Hsing et al. (1988) showed that under certain conditions the extremal index is the inverse of the mean cluster size. It follows that

$$\theta^{-1} = \lim_{n\to\infty} \sum_{j=1}^{\infty} j\pi_n(j, u_n, r_n)$$

Summarizing the above results, univariate EVT provides probabilistic tools to model the limiting distributions of normalized maxima and excesses over high thresholds. Regarding the parameter estimation of extreme value distributions it suffices to apply parametric estimation methods instead of nonparametric estimation methods which are less robust for small sample sizes.

2.3. Multivariate extremes

In this section, we study the limiting distributions of componentwise defined maxima of iid d-variate random vectors. A first problem that arises, when $d > 1$, is the lack of a natural definition of extreme values, since different concepts of ordering are possible. The difficulty for statistical applications of multivariate extreme value modes is due to the number of dimensions d increase considerably these do not reduce to a finite dimensional parametric family, so there is potential explosion in the class of models to be considered. The probabilistic limit theory of multivariate extremes is reasonably well established, and has been reviewed in the books of Resnick (1987) and Galambos (1975), but statistical theory is still in rapid development. Coles and Tawn (1991, 1994) have proposed methods based on multivariate extreme value distributions, which therefore generalize the classical approach to univariate extremes based on the limiting extreme value distributions, while Coles and Tawn (1999), proposed methods extending the threshold-exceedances approaches developed in Smith (1989).

The following definition of multivariate extreme values seems to be useful for several practical cases, as was explained by Resnick (1987). References for this section include Falk et al. (2004); Kotz and Nadarajah (2002); Galambos (1978).

2.3.1. Multivariate extreme value theory for iid sequences. Let $\mathbf{X}_\mathrm{i} = X_{i,1}, \ldots, X_{i,d}$ $i = 1, 2, \ldots, n$ be iid $d-$variate random vectors with join distribution function F and

$$\mathbf{M}_\mathrm{n} = (\mathrm{M}_{\mathrm{n},1}, \ldots, \mathrm{M}_{\mathrm{n,d}}) = \left(\max_{1\leq \mathrm{i} \leq \mathrm{n}} \mathrm{X}_{\mathrm{i},1}, \ldots, \max_{1\leq \mathrm{i} \leq \mathrm{n}} \mathrm{X}_{\mathrm{i,d}} \right)$$

is the vector of maxima of each component.

Before continuing we make some remarks about the notation: from now on relations and operations are taken componentwise. Thus for $\mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^\mathrm{d}$ we write $\mathbf{x} < \mathbf{y}$ and $\mathbf{ax} + \mathbf{b}$ to denote the relations $x_j < y_j$ for all $j = 1, \ldots, d$ and the vector $(a_1x_1 + b_1, \ldots, a_dx_d + b_d)$ respectively.

We are interested in the asymptotic behaviour of $\mathbf{M}_n$. Reasoning analogously as in the univariate case, we can say that we need to normalize $\mathbf{M}_n$. Thus, the aim of multivariate extreme value theory is to seek conditions on the distribution function F of $\mathbf{X}_i$ under which there are sequences $\mathbf{a}_n$ and $\mathbf{b}_n$ such that

$$\mathbb{P}\{\mathbf{M}_{\mathrm{n}} \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n\} = F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \to G(\mathbf{x}), \tag{2.3.1}$$

converges to a non-degenerate $d-$variate distribution function G as $n \to \infty$. If this holds for suitable choices of $\mathbf{a}_n > 0$ and $\mathbf{b}_n$, then we say G is a multivariate extreme value distribution (MEVD) and F is in the domain of attraction of G, $F \in MDA(G)$. In contrast with the univariate case, there is no finite-dimensional parametric family that covers the whole class.

In a multivariate context, a distribution function G is *max-stable* if for all $j = 1, \dots, d$ and $n > 2$ there are constants $\mathbf{a}_n > 0$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that

$$G^n(\mathbf{x}) = G(\mathbf{x}^{1/n}).$$

If the G_j are univariate max-stable distribution functions, then one can prove that

$$\prod_{j \leq d} G_j(x_j) \leq G(\mathbf{x}) \leq \min\left(G_1\left(x_1\right), \dots, G_d\left(x_d\right)\right)$$

for every max-stable distribution G with marginals G_j, where the right-hand side represents the case of totally dependent random variables and the left-hand side the case of independent random variables. Max-stable distributions form a subclass of max-infinitely divisible (max-id) distributions which is the class of all distribution functions G, such that for all $t > 0$, G^t is again a distribution function. For some basic facts about max-stability and max-infinite divisibility see Resnick (1987); Falk et al. (2004).

EXAMPLE 2.3.1. Consider a multivariate Gaussian distribution, with all univariate marginals equal to N(0, 1), and with all its correlations less than 1. Such a distribution is in the domain of attraction of the independence with univariate Gumbel marginals. Indeed, one has that

$$F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \to G(\mathbf{x}) = \prod_{j=1}^{d} \exp\left\{-e^{-x_j}\right\}.$$

The norming constants are respectively equal to $\mathbf{a}_n = (2\log n)^{-1/2}$ and $\mathbf{b}_n = b_n\mathbf{1}$, where $b_n = (2\log n)^{1/2} - 1/2\left(\log\log n + \log 4\pi\right) / \left(2\log n\right)^{-1/2}$, and $\mathbf{1} = (1, \dots, 1)$.

2.3.2. Characterization of multivariate extreme value distributions. In contrast to the univariate case, the multivariate extreme value distributions cannot be represented by a parametric family indexed by a finite-dimensional parameter vector, the class of dependence structures is simply too large. Instead the family of multivariate extreme value distributions can be indexed by a class of finite measures; exponent measures (Starica (1999); Heffernan and Resnick (2005); Resnick (2006); Balkema and Embrechts (2007)) or in another description by a class of dependence functions as the Pickands dependence functions (Kotz and Nadarajah (2002); Klüppelberg and Mayer (2006)) or Copulas functions (Joe (1997); Embrechts et al. (2003); Nelsen (2006)). However, all the representations are equivalent.

In order to give a characterization of max-stable distributions or equivalently of limit distributions for appropriately multivariate extreme value models, it is an enormous help to first standardize the problem so that G has specified marginals G_j. In doing so we do not only get simpler expressions but also, and this is even more important, we can separate dependence aspects from marginal distributions features. Marginals are handled using univariate Frèchet random variables, whereas the dependence structure needs some closer consideration.

Before we give a complete definition of multivariate extreme value theory, let us concentrate on some of these dependence representations.

THEOREM 2.3.2. *(Characterization of multivariate extreme value distributions)*
The following statements are equivalent:

- *G is a multivariate extreme value distribution with Φ_1 marginals.*
- *There exists a finite measure $\mathcal{S}$ on $\mathbb{S}^{d-1} = \left\{z \in \mathbb{R}^d : \|z\| = 1\right\}$, such that for each $\mathbf{x} \in \mathbb{R}^d$ one has that*

$$G(\mathbf{x}) = \exp\left\{-\int_{\mathbb{S}^{d-1}} \bigvee_{j=1}^{d} \left(\frac{a_i}{x_i}\right) \mathcal{S}(d\boldsymbol{a})\right\},$$

with $\int_{\mathbb{S}^{d-1}} \bigvee_{j=1}^{d} a_j \mathcal{S}(d\boldsymbol{a}) = 1$ for all $j = 1, \ldots, d$.
- *(Point process characterization)*
There exists $N(\cdot) = \sum_{k=1}^{\infty} \mathbb{I}_{(t_k, j_k)}(\cdot)$, where $N \sim PRM(dt \times d\mu)$ on $[0, \infty) \times \mathbb{C}$ and $\mathbb{C} = [0, \infty]^d \setminus \{\mathbf{0}\}$ with

$$\mu\left(\boldsymbol{y} \in \mathbb{C} : \|\boldsymbol{y}\| > r;\ \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \in A\right) = r^{-1}\mathcal{S}(A)$$

and S a finite measure such that for $\mathbf{x} \geq 0$

$$G(\mathbf{x}) = \mathbb{P}\left(\bigvee_{t_k \leq 1} j_k \leq \mathbf{x}\right) = \exp\left\{-\mu\left([\mathbf{0}, \mathbf{x}]^c\right)\right\}.$$

Recently the idea of multivariate extreme value results in terms of copula has taken place in the academic word. The idea behind the concept of copulas is to separate a multivariate distribution function into two parts, one describing the dependence structure and the other describing marginal behaviour, respectively.

Copulas are an extremely useful concept because this allows us to express dependence on a quantile scale, which is important for describing the dependence of extreme events. Moreover, they allow combining marginal models with different possible dependence models and therefore, to investigate the sensitivity of risk to the dependence specification. We begin with some definitions and theorems which will turn out to be useful to understand the ideas behind copulas.

DEFINITION 2.3.3. A d-dimensional copula is a distribution function on $[0,1]^d$ with standard uniform marginal distributions. Equivalently a copula is any function $C : [0,1]^d \to [0,1]$ which has the following three properties:

(1) $C(u_1, \ldots, u_d)$ is increasing in each component.
(2) $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$, $u_i \in [0,1]$.
(3) For all $(a_1, \ldots a_d)$,$(b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \leq b_i$ we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0, \tag{2.3.2}$$

where $u_{j1} = a_j$ and $x_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$.

These three properties characterize a copula. Since a copula is the distribution function of a random vector with uniform marginals, it is a continuous function. The following elegant theorem, due to Sklar, states that in the case of continuous marginals C is unique, then copulas in conjunction with univariate distribution functions may be used to construct multivariate distribution functions.

THEOREM 2.3.4. . *Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a a random vector with joint distribution function F and let $F_1, \ldots, F_d$ be the marginal distributions. Then, there exists a copula $C : [0,1]^d \to [0,1]$ such that the joint distribution can be written as*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{2.3.3}$$

If the marginals are continuous, then C is unique; otherwise C is uniquely determined on $Ran\, F_1 \times Ran\, F_2 \times \cdots \times Ran\, F_d$, where $Ran\, F_i$ denotes the range of F_i. Conversely, for a given copula C and marginals $F_{1,\ldots,}F_d$ we have that (2.3.3) defines a distribution with marginals F_i.

PROOF. See Nelsen (2006), pp. 18. □

This theorem motives the name of copula. In fact (2.3.3) means that C couples the marginals F_i to the joint distribution function F. One useful property of the copula of a distribution is its invariance under strictly increasing transformations of the marginals.

In the literature there are several types of copula as well as in the Pickad's dependence function, but there are three of particular importance, the comonotonicity copula or Frèchet-Hoeffding upper bound copula, which correspond to perfect dependence, the independence copula which is the copula of independent random variables and the countermonotonicity copula or Frèchet-Hoeffding lower bound copula, which implies perfect negative dependence. Following we introduced these intuitively.

We know that for any multivariate distribution F function with marginals $F_1, \ldots, F_d$ we may be given the following bounds.

$$\max\left\{\sum_{i=1}^{d} F_i(x_i) + 1 - d, 0\right\} \leq F(x) \leq \min\{F(x_1), \ldots F(x_d)\}.$$

By similar reasoning and with the help of Theorem 2.3.4 we can rewrite that last equation in terms of copulas. Thus, for every copula we have the bounds.

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \leq C(\mathbf{u}) \leq \min\{u_i, \ldots u_d\},$$

where the left side is the Frèchet -Hoeffding lower bound, which is only a copula for $d = 2$, and the right side function is Frèchet -Hoeffding upper bound copula. Perfect negative dependence extension to dimensions higher than two is not possible, because the Frèchet lower bound is not a proper distribution function and does not satisfy (2.3.2), which is necessarily to ensure that if a random vector $(U_1, \ldots, U_d)$ has a distribution function C. Then, $\mathbb{P}(a_1 \leq U_1 \leq b_1, \ldots, a_d \leq U_d \leq b_d)$ is non-negative, where a_i and b_i are defined as in Definition 2.3.3. These three fundamental copulas are displayed in Figure 2.3.1.

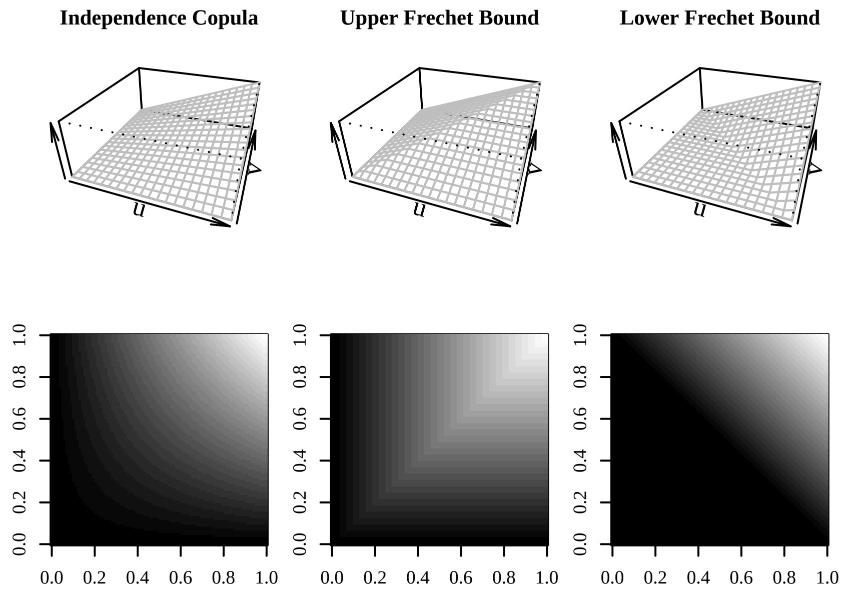

FIGURE 2.3.1. Perspective plots and heat maps contour of the three fundamental copulas, Independence copula, Upper Frèchet-Hoeffding upper bound copula, and the Frèchet-Hoeffding lower bound copula for $d = 2$. Darker colours correspond to stronger dependence.

In the case of multivariate extreme value theory the copulas are derived from the dependence structure of multivariate extreme value distributions, which provide the limit distributions of the component-wise maxima of d-dimensional random vectors, after a suitable normalization.

Let $\mathbf{X}_{\mathrm{i}} = X_{i,1}, \ldots, X_{i,d}$, $i = 1, 2, \ldots, n$ be iid $d-$variate random vectors with join distribution function F and $\mathbf{M}_{\mathrm{n}} = (\mathrm{M}_{\mathrm{n},1}, \ldots, \mathrm{M}_{\mathrm{n,d}}) = (\max_{1\leq \mathrm{i} \leq \mathrm{n}} \mathrm{X}_{\mathrm{i},1}, \ldots, \max_{1\leq \mathrm{i}\leq \mathrm{n}} \mathrm{X}_{\mathrm{i,d}})$ their component-wise maxima. We know from (2.3.1) that if a limiting distribution exists, then each univariate marginal of this distribution is an univariate extreme value distribution. Therefore, it can be written as follows

$$C_e \left(G\left(x_1\right), \ldots, G\left(x_d\right)\right),$$

where G is a extreme value distribution, and C_e is an d-extreme value copula. Joe (1990) shown that in particular the copulas satisfying the condition

$$C_e \left(\mathbf{u}^t\right) = C_e \left(\mathbf{u}\right)^t$$

for all $t > 0$ is a extreme value copulas. The following theorem makes the connexion between Pickand's dependence function and extreme value copulas.

THEOREM 2.3.5. *(Pickand's representation) The copula C_e is an Extreme value copula for G if and only if*

$$C_e \left(\mathbf{u}\right) = \exp \left\{ \left(\sum_{i=1}^{d} \ln u_i \right) A \left(\frac{\ln u_1}{\sum_{i=1}^{d} \ln u_i}, \ldots, \frac{\ln u_d}{\sum_{i=1}^{d} \ln u_i} \right) \right\}$$

where $A\left(v\right)$ is defined as Pickands dependence function.

Thus, we can extract the dependence function from the extreme value copula and vice versa by means of

$$A(v) = -\ln C(\exp(-w), \exp-(1-w)) \tag{2.3.4}$$

in the bivariate case.

The function $A(v)$ is commonly called the *Pickands dependence function* of the multivariate extreme value distribution. This characterizes the dependence structure for G. Note that, however, we can express G in terms of Pickands dependence function without the help of Copula function. Indeed, in Theorem 2.3.2 we have that $G(\mathbf{x}) = \exp\{-V(\mathbf{x})\}$ where

$$\begin{aligned} V(\mathbf{x}) &= V(x_1, \ldots, x_d) = \int_{\mathbb{S}^{d-1}} \bigvee_{j=1}^{d} \left(\frac{a_i}{x_i}\right) \mathcal{S}(d\boldsymbol{a}) \\ &= \left(\frac{1}{x_1} + \cdots + \frac{1}{x_d}\right) \int_{\mathbb{S}^{d-1}} \bigvee_{j=1}^{d} \left(a_1 \frac{x_1^{-1}}{\sum_{j=1}^{d} x_j^{-1}}, \ldots, a_d \frac{x_d^{-1}}{\sum_{j=1}^{d} x_j^{-1}},\right) \mathcal{S}(d\boldsymbol{a}) \\ &= \left(\sum_{j=1}^{d} x_j^{-1}\right) A\left(\frac{x_1^{-1}}{\sum_{j=1}^{d} x_1^{-1}}, \ldots, \frac{x_d^{-1}}{\sum_{j=1}^{d} x_j^{-1}}\right), \end{aligned}$$

with

$$A(v) = \int_{\mathbb{S}^{d-1}} \bigvee_{j=1}^{d} (v_j a_j) \mathcal{S}(d\boldsymbol{a}).$$

The V function is often called exponent measure and it is a homogeneous function of order -1. Contrary to the univariate case, there is an infinity of functions V for $d > 1$ as in the case of the copula function. It can be verified that for A has the following properties.

- $A(0) = A(1) = 1$;
- $-1 \leq A'(0) \leq 0$;
- $0 \leq A'(1) \leq 1$; $A''(v) \geq 0$ and $\max(v_1, \ldots, v_d) \leq A(v) \leq 1$ for all $v \in \mathbb{S}^{d-1}$
- $A(v) = 1$implies that the marginals are totally independent;
- $A(v) = \max(v_1, \ldots, v_d)$ implies that the marginals are totally dependent;
- A is a convex function.

Some multivariate extreme value models in terms of their Pickands dependence function or copulas are for instance, logistic (Galambos (1975)), negative Logistic (Joe (1990)), Hüsler-Reis (Falk et al. (2004)), Dirichlet (Coles and Tawn (1994)), asymmetric mixed model (Tawn (1988a)). For more detailed formulation and other results see Kotz and Nadarajah (2002).

Example 2.3.6. The easiest example of extreme value copula is the independent copula. Let G be a extreme value distribution with dependence function $A(v)$. We know that the independence case for the Pickands representation of a multivariate extreme value distribution correspond to $A(v) = 1$, then by

$$\begin{aligned} C(\mathbf{u}) &= G(\ln u_1, \ldots, \ln u_d) \\ &= \exp\left\{\left(\sum_{i=1}^{d} \ln u_i\right) A\left(\frac{\ln u_1}{\sum_{i=1}^{d} \ln u_i}, \ldots, \frac{\ln u_d}{\sum_{i=1}^{d} \ln u_i}\right)\right\} \\ &= \left(\prod_{i=1}^{d} u_i\right)^{A\left(\frac{\ln u_1}{\sum_{i=1}^{d} \ln u_i}, \ldots, \frac{\ln u_d}{\sum_{i=1}^{d} \ln u_i}\right)} \end{aligned}$$

and because $A(v) = 1$ the last equation yields the independence copula

$$C(\mathbf{u}) = \left(\prod_{i=1}^{d} u_i\right).$$

Exactly as for the maximum domain of attraction for the univariate case, we need to define this for the multivariate case. We now give this extension based in copulas.

THEOREM 2.3.7. *The copula C is said to belong to the domain of attraction of an extreme value copula C_e, written $C \in CDA(C_e)$, if and only if*

$$C(F_1(x_1), \ldots, F_d(x_d)) \in MDA(C_e(G_1(x_1), \ldots G_d(x_d)))$$

whenever $F_j(x_j)$ is continuous and belongs to $MDA(G_j)$, for $j \in \{1, \ldots, d\}$, and

$$\lim_{t \to \infty} C^t\left(\mathbf{u}^{1/t}\right) = C_e(\mathbf{u})$$

for $\mathbf{u} \in [0,1]^d$.

From the last theorem we can see that the determination of marginals $G_1, \ldots G_d$ depend only on the marginals $F_1, \ldots, F_d$. Then, for the estimation of its dependence structure are irrelevant. Hence that we need to define other concepts to establish whether a given copula is in the domain of attraction of some other give copula. The next definition is due to McNeil et al. (2005), pp. 315.

DEFINITION 2.3.8. Let C be a copula and C_e an extreme value copula, if $C \in CDA(C_e)$ then

$$\lim_{t \to 0} \frac{1 - C(1 - tx_1, \ldots, 1 - tx_d)}{t} = -\ln C_e(\exp(-x_1), \ldots, \exp(-x_1)).$$

The following result is of central relevance in multivariate extreme value theory. It states that the asymptotic behaviour of maxima of random vectors can be described by considering marginals and copulas separately. In other words, we assert that the copula C_e of H is only determined by the copula C of F, but it is in no way influenced by the marginals of F.

THEOREM 2.3.9. *Let $F(\mathbf{x}) = C(F(x_1), \ldots, F(x_d))$ for continuous marginal distribution functions $F_1, \ldots, F_d$ and some copula C. Let $H(\mathbf{x}) = C_e(H_1(x_1), \ldots, H_d(x_d))$ be an multivariate extreme value distribution with extreme value copula C_e. Then, $F \in MDA(H)$ if and only if $F_j \in MDA(H_j)$ for $j = 1, \ldots, d$ and*

$$\lim_{t \to \infty} C^t\left(u_1^{1/t}, \ldots u_d^{1/t}\right) = C_e(u_1, \ldots, u_d),$$

for $\mathbf{u} \in [0,1]^d$.

We briefly summarize the main ideas of how works the dependence in multivariate extreme value theory. For other types see Kotz and Nadarajah (2002); Falk et al. (2004) form the point of view of Pickands dependence and Joe (1997); Capéraà et al. (1997); Nelsen (2006) in the case of copulas.

Example 2.3.10. The logistic family of distributions have been among the most applied multivariate extreme value distribution in the literature because of its simplicity. The distribution function G of a bivariate logistic Model is defined by

$$G(x_1,x_2) = \exp\left[-\frac{1-\psi_1}{x_1} - \frac{1-\psi_2}{x_2} - \left\{\left(\frac{\psi_1}{x_1}\right)^q + \left(\frac{\psi_2}{x_2}\right)^q\right\}^{1/q}\right]$$

where $\psi_1, \psi_2 \in [0,1]$ and $q > 1$. The density of the measure $\mathcal{S}$ in Theorem 2.3.2 can be derived as follows

$$h(v) = \{(\psi_2 v)^q + (\psi_1 (1-v))^q\}^{1/q-2} \{v(1-v)\}^{q-2} (\psi_1\psi_2)^q (1-q).$$

We can show that the Logistic distribution has mass both in the interior and at the end points because $\mathcal{S}(\{0\}) = 1-\psi_2$ and $\mathcal{S}(\{1\}) = 1-\psi_1$. Independence is obtained when either $q \to 1^+$, $\psi_1 = 0$ or $\psi_2 = 0$ and total dependence correspond to $\psi_1 = \psi_2 = 1$ and the limit $q \to \infty$. The Pickands' dependence function is defined as

$$A(v) = (1-\psi_1)(1-v) + (1-\psi_2)\,v + \left\{(1-v)^{-q}\,\psi_1^{-q} + v^{-q}\psi_2^{-q}\right\}^{1/q}.$$

Then, we can create the extreme value copula by means of the equation (2.3.4), i.e., the copula

$$C_{q,\psi_1,\psi_2}(u_1,u_2) = u_1^{1-\psi_2} u_2^{1-\psi_1} \exp\left\{-\left((\psi_2 \ln u_1)^{-q} + (-\psi_1 \ln u_2)^{-q}\right)^{1/q}\right\}.$$

Special cases of this general formulation are the symmetric logistic distribution and the Gumbel distribution. The first case results from $\psi_1 = \psi_2 = 1$ having all its mass in the interior

$$G(x_1,x_2) = \exp\left\{-\left(x_1^{-q} + x_2^{-q}\right)\right\}^{1/q}.$$

The second case if $\psi_1 = \psi_2 = \alpha$ we have the Gumbel distribution

$$G(x_1,x_2) = \exp\left[-\frac{1-\alpha}{x_1} - \frac{1-\alpha}{x_2} - \alpha\left\{\left(\frac{1}{x_1}\right)^q + \left(\frac{1}{x_2}\right)^q\right\}^{1/q}\right].$$

A direct generalization of this logistic families is possible for the multivariate case. Let C be the a index variable over the set B, the class of all nonempty subsets of $\{1,\ldots,d\}$, let $B_1 = \{C \in B : |C| = 1\}$ and $B_{(j)} = \{b \in B : j \in C\}$, where $|C|$ denote the number of elements in the set C. The multivariate Logistic model is given by

$$G(x_1,\ldots x_d) = \exp\left[\sum_{C\in B}\left\{\sum_{j\in C}(\psi_{j,C}/x_j)^{q_c}\right\}^{1/q_c}\right], \tag{2.3.5}$$

where the dependence parameters $q_c \geq 1$ for all $C \in B \setminus B_1$, and the parameters $\psi_{j,C} = 0$ if $j \notin C$ else $0 \leq \psi_{j,C} \leq 1$ and $\sum_{C\in B}\psi_{j,C} = 1$ to ensure that the marginal distributions are Generalized extreme value distributions for $j = 1,\ldots,d$.

For the density estimation we refer to Kotz and Nadarajah (2002), pp. 121. Setting $\psi_{j,C} = 1$ and $q_C = q$ for all $j = 1, \ldots, d$, into (2.3.5) we obtain the symmetric logistic distribution

$$G\left(x_1, \ldots x_d\right) = \exp\left[\sum_{j=1}^{d} -\left(x_j^{-q}\right)^{1/q}\right].$$

Other special cases can be derived as limits of (2.3.5) as for example the Marshall and Olkin, and the MacFadden distribution , when $q_C \to \infty$ for all $C \in B$.

We simulate a trivariate sample of size 1000 with Frèchet marginals and a Gumbel dependence measure with levels of dependence $p = \{0.1, 0.5, 0.9\}$, i.e., near independence, weak dependence and strong dependence. The results are presented in Figure 2.3.2.

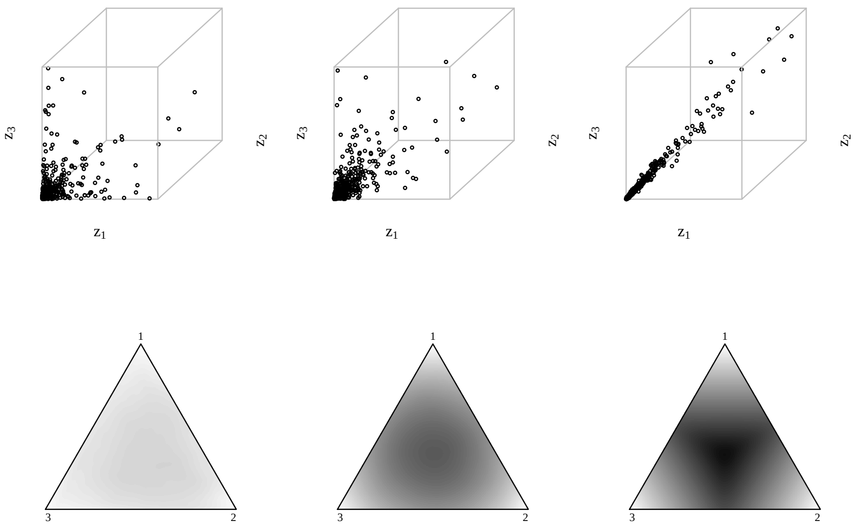

FIGURE 2.3.2. Upper panel: Simulation of a trivariate sample of Frèchet marginals with Logistic dependence measure, for $p = 0.9$ (left), $p = 0.5$ (middle), $p = 0.1$ (right). Lower panel: non-parametric estimation of dependence measure $A(v)$ for the simulations of the trivariate distributions in the Upper panel. The triangular border represents the constraint $\max_{j=\{1,\ldots,d\}} v_j \leq A\left(\mathbf{v}\right) \leq 1$ for all $v_j \in [0, 1]$. Darker colours correspond to smaller values of $A(v)$, and hence stronger dependence.

2.3.3. Multivariate extreme value theory for stationary sequence. Define $\mathbf{X}$ be a d-dimensional stationary sequence with a common distribution function $F(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, in the domain of attraction of a multivariate extreme value distribution H. This means that there exist a vector of constants $\mathbf{a}_n$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that for a high threshold $\mathbf{u}(\mathbf{x})$

$$F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) = F^n(\mathbf{u}(\mathbf{x})) \to H(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d. \tag{2.3.6}$$

If $\tilde{\mathbf{M}}_{\mathrm{n}} = \left\{\tilde{\mathbf{M}}_{\mathrm{n}1}, \ldots, \tilde{\mathbf{M}}_{\mathrm{n}d}\right\}$ denotes the vector of maxima corresponding to an iid sequence $\tilde{\mathbf{X}}$ having the same d- dimensional distribution function F, then (2.3.6) is equivalent to

$$\mathbb{P}(\tilde{\mathbf{M}}_{\mathrm{n}} \leq \mathbf{u}_n(\mathbf{x})) = \mathbb{P}\left(\bigcap_{j=1}^{d} (\tilde{M}_{nj} \leq a_{nj}x_j + b_{nj})\right) \to \tilde{H}(\mathbf{x}),$$

where $\tilde{H}$ is a distribution function with non-degenerate marginals.

For stationary sequence satisfying a general mixing assumption, it is known that the class of limiting distributions of $\mathbf{M}_{\mathrm{n}}$ is the same as for iid sequences. The following theorem uses a multivariate version of the $\triangle(u_n)$ condition, proposed by Nandagopalan (1994), in which the only difference is that the random variables and the thresholds have dimension d.

Theorem 2.3.11. *Let $\tilde{\mathbf{X}}$ satisfy the long range dependence condition$\triangle(\mathbf{u}_n)$ and suppose that*

$$\mathbb{P}(\mathbf{M}_{\mathrm{n}} \leq \mathbf{u}_n(\mathbf{x})) \xrightarrow{w} H(\mathbf{x}),$$

where H is a distribution function with non-degenerate components, then the distribution function H is also a multivariate extreme value distribution.

The relation between the multivariate extreme value distributions H and $\tilde{H}$ can be expressed by means of the multivariate extremal index function $\theta(\tau)$, $\tau \in \mathbb{R}_+^d$, introduced by Nandagopalan (1994).

Theorem 2.3.12. *(Multivariate extremal index). Let $\mathbf{X}$ satisfy the long range dependence condition$\triangle(\mathbf{u}_n)$. Suppose that $\mathbb{P}(\mathbf{M}_{\mathrm{n}} \leq \mathbf{u}_n(\mathbf{x})) \xrightarrow{w} H(\mathbf{x})$, and $\mathbb{P}(\tilde{\mathbf{M}}_{\mathrm{n}} \leq \mathbf{u}_n(\mathbf{x})) \xrightarrow{w} \tilde{H}(\mathbf{x})$. if for all $\tau \in \mathbb{R}$ there exists $\mathbf{u}_n^{(\tau)}$. Then, the multivariate extremal index of $\mathbf{X}$ is defined by the relation $H(\mathbf{x}) = \tilde{H}^{\theta(\tau)}(\mathbf{x})$, with $\theta(\tau) \in [0, 1]$, for all $\tau \in \mathbb{R}$.*

Just as in one dimension it is the key parameter relating the extreme value properties of a stationary process to those of independent random vectors from the same d-dimensional marginal distribution. However, unlike the one dimensional case, it is not a constant for the whole process, but instead dependences on the vector τ. Some elementary properties include:

(1) $0 \leq \theta(\tau) \leq 1$ for all τ.
(2) For each $j = 1, \ldots, d$, $\mathbf{X}_{ij}$ has extremal index $\theta_j = \lim_{\tau_{i_{i\neq j}} \to 0+} \theta(\tau_1, \ldots, \tau_d)$.
(3) $\theta(c\tau) = \theta(\tau)$ for all $c > 0$. (Theorem 1.1 of Nandagopalan (1994))

A version of the distribution of cluster in d-dimensions can be defined as follow.

Definition 2.3.13. (Cluster size probability) Consider a sequence of integers k_n and r_n, and a random matrix $\mathcal{M}_{r_n}$ with dimensions $r_n \times d$, consisting of $1'$s and $0'$s accordingly to $X_{ij} > u_{nj}^{(\tau_j)}$ or not. That is, $\mathcal{M}_{r_n}(i.j) = \mathbb{I}\left\{X_{ij} > u_{nj}^{(\tau_j)}\right\}$, $i = 1, \ldots, r_n$, $j = 1, \ldots, d$. If $\mathbf{X}$ satisfies the long range dependence condition$\triangle(\mathbf{u}_n)$, then the probability distribution

$$\pi_n^{(\tau)}(k_n) = \mathbb{P}\left\{\sum_{i=1}^{r_n} \mathbb{I}\left\{\mathcal{M}_{r_n}^{(i)} \neq 0\right\} = k \,\middle|\, \sum_{i=1}^{r_n} \mathbb{I}\left\{\mathcal{M}_{r_n}^{(i)} \neq 0\right\} > 0\right\},\ k \geq 1,$$

satisfies

$$\theta^{-1}(\tau) = \lim_{n\to\infty} \sum_{k\geq 1}^{\infty} k\pi_n^{(\tau)}(k).$$

Therefore, for weakly dependent sequences $\mathbf{X}$, the existence of $\theta(\tau) \in [0,1]$ indicates the presence of clusters of exceedances $\left\{\mathbf{X}_i \nleq \mathbf{u}_n^{(\tau)}\right\}$, $i = 1, \ldots, r_n$. Moreover, for each dimension, we can obtain information about $\theta(\tau)$ on the presence of marginal cluster of exceedances $\left\{\mathbf{X}_j \nleq \mathbf{u}_{nj}^{(\tau_j)}\right\}$, i.e., $X_{ij} > u_{nj}$ for some $j = 1, \ldots, d$, as a consequence of the definition of the multivariate extremal index.

An alternative characterization of the multivariate extremal index is given by Smith and Weissman (1996), but this will be introduced in the next chapter.

2.4. Extreme measures of dependence

In the above sections we showed how copulas and Pickands dependence functions give a complete description of the whole dependence structure of a multivariate random vector. However, in many situations the type of dependence is unknown, therefore it should be chosen from a wide class of extreme dependence functions. These dependence functions should be as flexible as possible in the range of asymptotic dependence behaviour which they can describe. For this reason, several measure of extreme dependence has been introduced as a common yardstick for the true type of asymptotic dependence behaviour.

In this section we only focus on extremes by introducing some scalar measures of dependence. These measures relate to the (asymptotic) behaviour of the tails of a distribution. Being scalar measures they will not be able to provide a complete picture of a dependence structure but they will turn out to be useful for distinguishing between different kinds of possible dependence behaviours. These measures can be expressed in terms of copulas in contrast to other dependence measures such as linear correlation, they are not influenced by the marginal distributions of the random vector (see for instance Embrechts et al. (2002)). These measures quantify in a particular sense the probability of one random variable being extreme, given that the other one is extreme, i.e., they are asymptotic dependent. Thus, we do not consider here the asymptotic independent case. For different measures in the asymptotic independent case we refer to Ledford and Tawn (1996); Coles and Tawn (1999); Heffernan and Resnick (2005).

The first definition is the most common approach to tail dependence.

Definition 2.4.1. (Joe (1997)) Let (X_1, X_2) be a random pair with joint cumulative distribution function F and marginals F_1 and F_2. The coefficient

$$\lambda_u := \lim_{u \to 1^-} \mathbb{P}\left(F_2\left(X_2\right) > u \mid F_1\left(X_1\right) > u\right)$$

is called the upper tail dependence coefficient ($UTDC$), provided the limit $\lambda_u \in [0,1]$ exists. We say that the pair (X_1, X_2) is upper tail dependent if $\lambda_u > 0$ and upper tail independent if $\lambda_u = 0$.

Analogously, for the lower tail dependence coefficient ($LTDC$) we obtain

$$\lambda_l := \lim_{u \to 0^+} \mathbb{P}\left(F_2\left(X_2\right) \leq u \mid F_1\left(X_1\right) \leq u\right).$$

Thus, the tail dependence coefficients correspond to the probability that one marginal exceeds a high or low threshold u under the condition that the other marginal exceeds a high or low threshold respectively. The next properties are given for the upper tail copula only and we written only λ instead λ_u. This concept can also be defined via the notion of copula.

DEFINITION 2.4.2. Let C be the copula of a d-variate distribution , then the tail dependence coefficient of this distribution can be defined by means of of its bivariate marginals, i.e.

$$\begin{aligned} \lambda_{i,j} : &= \lim_{u \to 1^-} \mathbb{P}\left(F_j\left(X_j\right) > u \mid F_i\left(X_i\right) > u\right) \\ &= \lim_{u \to 1^-} \frac{\overline{C}_{i,j}\left(1-u, 1-u\right)}{1-u} \\ &= \lim_{u \to 0^+} \frac{\overline{C}_{i,j}\left(u, u\right)}{u} \end{aligned} \tag{2.4.1}$$

for all $i \neq j \in \{1, \ldots, d\}$, where $\overline{C}_{i,j}\left(1-t_i, 1-t_j\right) = 1 - t_i - t_j + C_{i,j}\left(t_i, t_j\right)$ denotes the survival copula of C.

Thus for a d-variate distribution we hope to calculate $d\left(d-1\right)/2$ upper or lower tail dependence coefficients. Since the tail dependence coefficients ($TDCs$) are determined by the copula of a distribution, then many copulas features are transferred to TDC, for example TDC are invariant under strictly increasing transformations of marginals.

Other interesting relation between TDC, Pickands dependence function and extreme value copula can be derived as follows. The first result is that if the copula $C \in CDA\left(C_e\right)$, then they have the same TDC. The second result is that the TDC is relationed with the Pickands dependence function by $\lambda = 2\left(1 - A\left(1/2\right)\right)$. We resume these results in the following theorem.

THEOREM 2.4.3. *(McNeil et al. (2005)) Let C be a bivariate copula with tail dependence coefficient $\lambda \in [0,1]$ and $C \in CDA\left(C_e\right)$ for some extreme value copula C_e. Then, λ is also the upper tail dependence coefficient of C_e. As consequence λ is related to Pickands dependence as $\lambda = 2\left(1 - A\left(1/2\right)\right)$.*

Other measure of extreme dependence is the extremal dependence coefficient (EDC) introduced by Frahm (2006).

DEFINITION 2.4.4. The extremal dependence coefficient of a d-dimensional random vector $\mathbf{X}$ with joint cumulative distribution function F and marginal distribution functions $F_1, \ldots, F_d$ is defined for the upper tail as

$$\varepsilon := \lim_{t \nearrow 1} \mathbb{P}\left(F_{\min} > t \mid F_{\max} > t\right),$$

where $F_{\max} := \max\left\{F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right\}$ and $F_{\min} := \min\left\{F_1\left(x_1\right), \ldots, F_d\left(x_d\right)\right\}$.

Thus the EDC can be interpreted as the probability that the best case of $\mathbf{X}$ leads to the worst case of $\mathbf{X}$. Whenever the EDC is positive the elements of $\mathbf{X}$ will be called extremal dependent.

Smith (1990) introduced a similar measure called the *extremal coefficient*, which is defined in term of some positive measure $\mathcal{H}$ on the d- dimensional simplex that satisfies

$$\int x_j d\mathcal{H}\left(x_1, \ldots, x_d\right) = 1 \tag{2.4.2}$$

for $j = 1, \ldots, d$.

DEFINITION 2.4.5. Let F be a multivariate extreme value distribution with unit Frèchet marginals

$$F(w_1, \ldots, w_d) = \exp\left(-\int \max\left\{\frac{x_1}{w_1}, \ldots, \frac{x_d}{w_d}\right\} d\mathcal{H}(x_1, \ldots, x_d)\right)$$

with $\mathcal{H}$ satisfying (2.4.2). The extremal coefficient is defined as the constant

$$\epsilon = \int \max_{j \in D} x_j d\mathcal{H}(x_1, \ldots, x_d) \tag{2.4.3}$$

In terms of extreme value copula that is

$$C^*(t, \ldots, t) = t^{\epsilon} \tag{2.4.4}$$

Moreover, Joe (1993) noticed, that the tail dependence coefficient can be expressed as $\lambda := 2 - \epsilon$. In the same form, the EDC can be related with other measures likes the extremal coefficient in the bivariate case as $\varepsilon_G = (2 - \epsilon)/\epsilon$.

2.5. Relation among the extreme measures of dependence in the stationary case

The main contribution of this section is to extend the extreme measures of dependence for iid random vectors to random vectors of stationary sequences. Schlather and Tawn (2002) showed that for a d-dimensional distribution exists 2^d distinct extremal coefficient of different orders, which cannot take any arbitrary values. Further, they constructed bounds that higher order extremal coefficients need to satisfy to be consistent with lower order extremal coefficients. In other way, Martins and Ferreira (2005) showed that the extremal coefficient does not measure correctly the dependence in the limiting distribution of maxima in presence of cluster of extremes in a stationary process and proposed an alternative definition in order to cover this drawback.

We extend these results to the other measures of extreme dependence and relate it with other measures of extreme dependence. In particular, we will present some bounds for the EDC for the three dimensional iid case. Then, the EDC will be extended to the stationary case. We show that, the EDC for absolutely dependent stationary sequence is completely determined by the extremal index of the marginals. Finally, we illustrate some results with examples.

Note that the EDC is a copula property, then if the copula of $\mathbf{X}$ belongs to a copula domain of attraction (CDA) of C_e, it is possible to express the definition of EDC in terms of its extreme value copula C_e.

THEOREM 2.5.1. *Let $\mathbf{X}$ be a d-dimensional iid random vector with joint distribution function $F = C(F_1(x_1), \ldots, F_d(x_d))$ for continuous marginal distribution functions $F_1, \ldots, F_d$. Further, let $\mathbf{M}_n := \max\{X_{n1}, \ldots, X_{nd}\}$ be denote as a d-dimensional vector of maxima. If there exist $\mathbf{a}_n > 0$, $\mathbf{b}_n \in \mathbb{R}^d$ when $n \to \infty$, such that $\mathbb{P}(\mathbf{M}_n \leq \mathbf{u}_n(x)) \to G(x)$, $x \in \mathbb{R}^d$, where $\mathbf{u}_n(x) = \mathbf{a}_n \mathbf{x}_n + \mathbf{b}_n$ and $G(x) = C_e(G_1(x_1), \ldots, G_d(x_d))$ is a multivariate extreme value distribution with non-degenerate marginals, then the EDC of F is also the EDC of G.*

2.5.1. Some sharp bounds for the EDC. In the definition of the EDC we shown that the EDC for bivariate case can be formulated in terms of the ϵ_G. Now we generalized this result to d-dimensions.

Proposition 2.5.2. *Let* $\mathbf{X}$ *be a d-dimensional iid random vector following a multivariate extreme value distribution with extremal coefficient* ϵ*, where* $D = \{1, \ldots, d\}$*. Define further* ϵ_{G_K} *as the extremal coefficient of a subset* $K \subseteq D$ *the marginals. Then, the EDC can be defined as*

$$\varepsilon_G = \frac{\sum_{K \subseteq D} (-1)^{|K|+1} \epsilon_{G_K}}{\epsilon_G}, \tag{2.5.1}$$

where $D := \{1, \ldots, d\}$*,* $\mathbb{I}_{\{\cdot\}}$ *is the indicator function and* $\epsilon_G = \epsilon_{G_D}$ *for ease of notation.. Moreover,* K *is a complete subset of all distinct combinations of marginals, i.e.,* 2^d *.*

Based on a result of Deheuvels (1983), Schlather and Tawn (2002) introduced a simple class of extreme value distributions that allows for a 1-1 mapping to the complete sets of extremal coefficients. In particular, if the extremal coefficients ϵ_K for the subset $K \in 2^D$ are given and the set is self-consistent, which signifies that $\epsilon_\emptyset = 0$ and

$$\sum_{K \in 2^D \setminus \{\emptyset\}, K \supseteq L} (-1)^{|K \setminus L|+1} \epsilon_K \geq 0 \quad \forall L \in 2^D \setminus \{D\}. \tag{2.5.2}$$

then ϵ is bounded by the sharp bounds

$$\max_{\substack{L \in 2^D \setminus \{D\} \\ |M \setminus L| \bmod 2 = 1}} \sum_{\substack{K \in 2^D \setminus \{\emptyset, D\} \\ K \supseteq L}} (-1)^{|K \setminus L|} \epsilon_K \;\leq\; \epsilon \;\leq\; \min_{\substack{L \in 2^D \setminus \{D\} \\ |M \setminus L| \bmod 2 = 0}} \sum_{\substack{K \in 2^D \setminus \{\emptyset, D\} \\ K \supseteq L}} (-1)^{|K \setminus L|+1} \epsilon_K. \tag{2.5.3}$$

This result allows the next sharp bounds for the EDC in the three dimensional case.

Proposition 2.5.3. *Let* $\mathbf{X}$ *be a d-dimensional iid random vector following a MVED in the* $MDA\,(G)$ *with a complete set of extremal coefficients, then,*
for d=3

(1) $\max^*_{i,j=1,2,3} \left\{0, (\varepsilon_{ij} + 1) \left[\frac{3}{2} - \sum_{k=1,2,3;\, k \neq j} \frac{1}{1+\varepsilon_{ik}}\right]\right\} \leq \varepsilon_{123} \leq \dfrac{\min^*_{i,j=1,2,3}\left(\frac{2\varepsilon_{ij}}{1+\varepsilon_{ij}}\right)}{\min_{i=1,2,3}\left(-1+\sum_{k=1,2,3;\, k \neq i} \frac{2}{1+\varepsilon_{ij}}\right)}$
where the sign "$*$" and "$\sum^*$" are defined as the combination and summation of indices whose members are pair-wise different.

(2) Additionally if two EDCs are known then the third has the next bounds
$\max\left\{0, \frac{1-\varepsilon_{ij}-\varepsilon_{ik}-3\varepsilon_{ik}\varepsilon_{ij}}{2(1+\varepsilon_{ik})(1+\varepsilon_{ij})}\right\} \leq \varepsilon_{jk} \leq 1 + \frac{(1+\varepsilon_{ij})(1+\varepsilon_{ik})-2|\varepsilon_{ik}-\varepsilon_{ij}|}{(1+\varepsilon_{ij})(1+\varepsilon_{ik})+2|\varepsilon_{ik}-\varepsilon_{ij}|}$.

(3) In terms of bivariate tail dependence parameters
$\frac{\max\{0, \lambda_{ij}+\lambda_{ki}-1\}}{\min\{2, 3-\lambda_{ij}-\lambda_{ki}\}} \leq \varepsilon_{jk} \leq \frac{1-|\lambda_{ki}-\lambda_{ij}|}{1+|\lambda_{ki}-\lambda_{ij}|}$.

Other extensions to higher dimensions are possible following the inequality (2.5.3). However, the results yield to a very difficult optimization problem to solve.

2.5.2. The extremal dependence coefficient of stationary sequences. In this section we suppose that the d-dimensional stationary sequence $\mathbf{X}$ with joint distribution function $F \in MDA(H)$ satisfies the long-range dependence conditions $\triangle(u_n)$ of Nandagopalan (1994) for $u_n > 0$ and $\hat{\mathbf{X}}$ is an associated iid sequence of $\mathbf{X}$ with joint distribution function $\hat{F} \in MDA(G)$ as in Theorem 2.3.12. A particularly interesting result is the relation between the multivariate extremal index $\theta(\tau)$ and the extreme value copulas of H and G.

PROPOSITION 2.5.4. *Let* $\mathbf{X}$ *be a* d*-dimensional stationary sequence with* $\theta(\tau)$, $\tau \in \mathbb{R}^d$, C_G^* *and* C_H^* *the extreme value copulas of the multivariate extreme value distributions* G *and* H *respectively, then*

$$C_H^*(t_1, \ldots, t_d) = C_G^*\left(t_1^{\theta_1}, \ldots, t_d^{\theta_d}\right) \tag{2.5.4}$$

Now, we are in condition to give the definition of the EDC for stationary sequence.

DEFINITION 2.5.5. *(EDC for stationary sequence)* Let $\mathbf{X}$ be a d- dimensional stationary sequence with joint distribution function $F \in MDA(H)$ and multivariate extremal index $\theta(\tau)$, $\tau \in \mathbb{R}^d$.

Further, let $F_{\max} := \max\{F_1(x_1), \ldots, F_d(x_d)\}$ and $F_{\min} := \min\{F_1(x_1), \ldots, F_d(x_d)\}$. Then, the EDC of $\mathbf{X}$ is defined as

$$\begin{aligned} \varepsilon &:= \lim_{t \nearrow 1} \mathbb{P}(F_{\min} > t \mid F_{\max} > t) \\ &= \lim_{t \nearrow 1} \frac{\mathbb{P}(F_1(x_1) > t, \ldots F_d(x_d) > t)}{1 - \mathbb{P}(F_1(x_1) \le t, \ldots F_d(x_d) \le t)}, \end{aligned}$$

provided the corresponding limits exist. In terms of copulas we have

$$\varepsilon_H := \lim_{t \nearrow 1} \frac{\overline{C_H^*}(1-t, \ldots, 1-t)}{1 - C_H^*(t, \ldots, t)}$$

where $\overline{C}$. is the survival copula belonging to C defined as

$$u \longmapsto \overline{C}(u) := \sum_{I \subset D} (-1)^{|I|} C\left((1-u_1)^{\mathbb{I}_{\{1 \in I\}}}, \ldots, (1-u_d)^{\mathbb{I}_{\{d \in I\}}}\right), \tag{2.5.5}$$

where $u \in [0,1]^d$, $D := \{1, \ldots, d\}$ and $\mathbb{I}_{\{\cdot\}}$is the indicator function.

The last result shows us that EDC is affected by the multivariate extremal index in the stationary case. Martins and Ferreira (2005) observed this problem for the extremal coefficient and gave a new definition for the extremal coefficient ϵ under cluster in the maxima.

DEFINITION 2.5.6. (Martins and Ferreira (2005)) Let $\mathbf{X}$ be a d-dimensional stationary sequence with joint distribution function $F \in MDA(H)$ and multivariate extremal index θ. The extremal coefficient of $H = G^\theta$ is the constant ϵ_H such that

$$H(\mathbf{x}) = G^{\theta(\mathbf{1})}(\mathbf{x}) = G_1^{\epsilon_H}(x), \, x \in \mathbb{R},$$

where $\theta(\mathbf{1}) = \theta(1, \ldots, 1)$.

Thus, the ϵ_H can be interpreted as the number of independent marginals involved in a d-dimensional multivariate extreme value distribution under the presence of cluster of extremes in the maxima, and takes values in $\left[\max_{j=1,\ldots,d}\theta_j, \sum_{j=1}^{d}\theta_j\right]$ for completely dependent and independent random variables respectively. See Proposition 2.1 in Martins and Ferreira (2005) for other interesting properties of ϵ_H.

Notice that when $\theta(\tau) = 1$, for all $\tau \in \mathbb{R}_+^d$, the Definition 2.5.5 of the EDC is the originally proposed for iid random variables by Frahm (2006).

2.5.3. General properties of the EDC of stationary sequences. First, we concentrate on the bivariate case to find relations between the EDC and other measures of extreme dependence.

PROPOSITION 2.5.7. *Let ϵ_H be the extremal coefficient of the limit distribution of the maxima of a d-dimensional stationary sequence $\mathbf{X}$ with joint distribution function $F \in MDA(H)$ and ε_H the corresponding EDC. Then,*

$$\varepsilon_H = \frac{\theta_1 + \theta_2 - \epsilon_H}{\epsilon_H} = \frac{\lambda_H}{\theta_1 + \theta_2 - \lambda_h}. \tag{2.5.6}$$

Thus, we can see that in the bivariate case the EDC ε_H has close relation to other extreme measures of dependence. Notice that the EDC for stationary sequence is always smaller than EDC ε_G for iid random vectors. For this is enough to rewrite (2.5.6) as $\varepsilon_H = \varepsilon_G - \frac{2\theta(1,1) - (\theta_1+\theta_2)}{\epsilon_G \theta(1,1)}$, which produces the next bounds for $\varepsilon_H \in [0, \min(\theta_1, \theta_2) / \max(\theta_1, \theta_2)]$.

We resume some direct results for the bivariate case in the next corollary.

COROLLARY 2.5.8. *Let $\mathbf{X}$ be a bivariate stationary sequence with joint distribution function $F \in MDA(H)$ and multivariate extremal index θ, where $\hat{\mathbf{X}}$ is the associated iid random vector of $\mathbf{X}$ with joint distribution function $\hat{F} \in MDA(G)$. Then,*

(1) $\varepsilon_H \in [0, \min(\theta_1, \theta_2) / \max(\theta_1, \theta_2)]$ and $\varepsilon_H \leq \varepsilon_G$.
(2) If $\theta_1 + \theta_2 \geq 1 + \lambda_H$ then $\varepsilon_H \geq \lambda_H$, else $\varepsilon_H < \lambda_H$.
(3) If $\epsilon_H = \max\{\theta_1, \theta_2\}$, then $\lambda_H = \min\{\theta_1, \theta_2\}$.

A special care must be taken when the bivariate asymptotic distribution of the maximum is completely dependent as the next example illustrates.

EXAMPLE 2.5.9. Let ψ_i be a sequence of iid Frèchet marginals and define the next fist-order ARMAX processes $Y_i = \max\{\beta Y_{i-1}, \psi_i\}$ and $X_i = \max\{\alpha X_{i-1}, \psi_i\}$ with $\alpha, \beta \in [0, 1)$ and $\alpha > \beta$. The bivariate Process $\mathbb{P}(X_i > u_n, Y_i > u_n)$ has bivariate Frèchet marginals with absolute dependent marginals and extremal indices $\theta_x = 1 - \alpha$, $\theta_y = 1 - \beta$ and $\theta_{xy} = \max\{1 - \alpha, 1 - \beta\}$. Whereas the associated bivariate iid random vector $\mathbb{P}\left(\hat{X}_i > u_n, \hat{Y}_i > u_n\right)$ has also completely dependent marginals with $\epsilon_G = 1$, $\lambda_G = 1$ and $\varepsilon_G = 1$.

By Proposition 2.5.7 the EDC $\varepsilon_H = 1 - \alpha/1 - \beta$. Thus, for example for $\alpha \uparrow 1$ and $\beta \downarrow 0$ the $\lambda_H = 1 - \alpha$, $\varepsilon_H \searrow 0$. Then, the value of EDC ε_H has not an easy interpretation as grade of asymptotic independence in a stationary sequence.

Now we want to know if we can extend this result to d-dimensional stationary sequences.

PROPOSITION 2.5.10. *Let $\mathbf{X}$ be a d- dimensional stationary sequence with joint distribution function $F \in MDA(H)$ and multivariate extremal index θ. Then,*

(1) If H has totally dependent marginals then $\varepsilon_H = \frac{\min_{1\leq j\leq d}\theta_j}{\max_{1\leq j\leq d}\theta_j}$.

(2) If H has totally dependent marginals and $\theta_1 = \cdots = \theta_d$ then $\varepsilon_H = 1$.
(3) If H has independent marginals then $\epsilon_H = \sum_{j=1}^{d} \theta_j$ and $\varepsilon_H = 0$.
(4) In the three dimensional case if $\varepsilon_{H_{ij}}$, $\varepsilon_{H_{ik}}$, θ_i, θ_j and θ_k are known, then $\varepsilon_{H_{jk}}$ has the next sharp bounds.
$\max\left\{0, \frac{\theta_i+\theta_k\varepsilon_{H_{ik}}+\theta_j\varepsilon_{H_{ij}}+\varepsilon_{ik}\varepsilon_{ij}(\theta_j+\theta_k-\theta_i)}{\sum_{l=\{i,j,k\}}\theta_l+\theta_j\varepsilon_{H_{ik}}+\theta_k\varepsilon_{H_{ij}}-\theta_i\varepsilon_{H_{ij}}\varepsilon_{H_{ik}}}\right\} \leq \varepsilon_{H_{jk}} \leq \min\left\{\widetilde{\varepsilon}_{jk}, \widetilde{\varepsilon}_{kj}\right\}$ where $\widetilde{\varepsilon}_{jk} := \frac{\theta_k+\varepsilon_{H_{ij}}\sum_{l=\{i,j,k\}}\theta_l+\varepsilon_{H_{ik}}(\theta_j\varepsilon_{ij}-\theta_i)}{\theta_j+\varepsilon_{H_{ik}}\sum_{l=\{i,j,k\}}\theta_l+\varepsilon_{H_{ij}}(\theta_k\varepsilon_{ik}-\theta_i)}$.

A direct result is the next Corollary.

Corollary 2.5.11. *Let* $\mathbf{X}$ *be a* d*-dimensional stationary sequence with joint distribution function* $F \in MDA(H)$ *and multivariate extremal index* $\theta(\tau) \in [0,1]^d$ *with absolutely dependent marginals. Subsequently, there are* $(d-1)$*-dimensional iid random vectors in* $\mathbf{X}$*, and only* 1*-dimensional stationary sequence in* X^* *with extremal index* θ^**. Then, the EDC* $\varepsilon_H = \theta^*$.

Frahm (2006) showed that if ε_X is the EDC of a d-dimensional vector X and $\overline{X}$ is a $(d-1)$-dimensional subset vector of X then $\varepsilon_{\overline{X}} \geq \varepsilon_X$. Loosely speaking, this could be interpreted as a sort of diversification effect at the extremes whether one adds a random component to a given random vector.

Unfortunately, it is not necessary true in the stationary case. To display this idea we observe the next example.

Example 2.5.12. (EDC of Multivariate Maxima of Moving Maxima (M_4))

Smith and Weissman (1996) introduced M_4 processes as an approximation to a large class of max-stable processes and hence to extremes of multivariate stationary processes. Let a_{lkj} $l \in \mathbb{N}$, $k \in \mathbb{Z}$, $j = \{1, \ldots, d\}$ be a triple sequence of non negative constants satisfying $\sum_l \sum_k a_{lkj} = 1$, for each j and let ψ_{lt}, $l \in \mathbb{N}$, $t \in \mathbb{Z}$ be a double sequence of iid unit Frèchet random variables. Then, M_4 process is defined by

$$X_{tj} = \max_{l\in\mathbb{N}} \max_{k\in\mathbb{Z}} a_{lkj}\psi_{l,t-k},$$

for $t\in\mathbb{Z}$, $j = \{1, \ldots, d\}$. Let $\mathbf{u}_n = (u_{n1}, \ldots, u_{nd}) = (n/\tau_1, \ldots, n/\tau_d)$ be the vector of normalized levels of $\mathbf{X}$. Subsequently, the distribution function of the maxima of this process is

$$\begin{aligned}
\mathbb{P}(M_n \leq \mathbf{u}_n) &= \mathbb{P}(X_{tj} \leq u_{tj},\ 1 \leq t \leq r,\ 1 \leq j \leq d) \\
&= \mathbb{P}\left(\psi_{l,t-k} \leq \frac{u_{tj}}{a_{lkj}},\ l \in \mathbb{N},\ k \in \mathbb{Z},\ 1 \leq t \leq r,\ 1 \leq j \leq d\right) \\
&= \mathbb{P}\left(\psi_{l,m} \leq \min_{1-m\leq k\leq r-m} \min_{1\leq j\leq d} \frac{u_{m+k,j}}{a_{lkj}},\ l \in \mathbb{N},\ m \in \mathbb{Z}\right) \\
&= \prod_{l\in\mathbb{N}} \prod_{m\in\mathbb{Z}} \mathbb{P}\left(\psi_{l,m} \leq \min_{1-m\leq k\leq r-m} \min_{1\leq j\leq d} \frac{u_{m+k,j}}{a_{lkj}}\right) \\
&= \prod_{l\in\mathbb{N}} \prod_{m\in\mathbb{Z}} C_H^*\left(\exp\left(\min_{1-m\leq k\leq r-m} \frac{a_{lk1}\tau_1}{n}\right), \ldots, \exp\left(\min_{1-m\leq k\leq r-m} \frac{a_{lkd}\tau_d}{n}\right)\right),
\end{aligned}$$

where $C_H^*(t) = \min(t_1, \ldots, t_d)$.

The extremal index for each marginal Maxima of Moving maxima is $\theta_j = \sum_{l\in\mathbb{N}} \max_{1-m\leq k\leq r-m} a_{lkj}$, $m \in \mathbb{Z}$. With the results obtained in Proposition 2.5.10 and the expression of the extremal index for each marginal, it is possible to derive directly the value for the EDC as follows

$$\varepsilon_H = \frac{\min_{1\leq j\leq d} \sum_{l\in\mathbb{N}} \max_{1-m\leq k\leq r-m} a_{lkj}}{\max_{1\leq j\leq d} \sum_{l\in\mathbb{N}} \max_{1-m\leq k\leq r-m} a_{lkj}}.$$

Hence, it could be possible to take, or add, a random component and not experiment any changes in the value of the EDC of this process, as long as one does it between the minimum and the maximum of the extremal indices of the marginals. Note that one could even remove the vector with the maximum extremal index to improve the value of the EDC ε_H, and one obtains consequently$\varepsilon_{\overline{X}} < \varepsilon_X$.

2.6. Conclusions

In this chapter we have presented the main concepts in extreme value theory for the iid and non iid case. The way in which we have presented the theory relied on the assumption that the random variables are independently distributed or that some mixing conditions hold in the non iid case. The extremal index, as measure of cluster at the extreme, has been highlighted. Apart from time dependence, the cross sectional dependence has also been introduced through copulas and Pickands dependence functions.

In Section 2.5 we investigated different measures of extreme dependence and established their equivalences. New relations in the stationary case were also derived. By observing the limit distribution of the maxima of the stationary sequence and the behaviour of the cluster extremes one can conclude, that the dependence at the extremes for different measures as the EDC can take very different values, when we work only with the associated iid random vector instead of the true stationary sequence.

2.A. Demonstrations

PROOF. Theorem 2.4.3: from (2.4.1) we have

$$\lambda = \lim_{u\to 1^-} \frac{\overline{C}(1-u, 1-u)}{1-u} = 2 - \lim_{u\to 1^-} \frac{C(u,u)}{1-u}.$$

Now using the asymptotic identity $\ln x \sim x - 1$ as $x \to 1$ and Definition 2.3.8 we obtained

$$\begin{aligned}
\lim_{u\to 1^-} \frac{1 - C_e(u,u)}{1-u} &= \lim_{u\to 1^-} \frac{\ln C_e(u,u)}{\ln u} \\
&= \lim_{u\to 1^-} \lim_{t\to 0^+} \frac{1 - \ln C\left(u^t, u^t\right)}{-t \ln u} \\
&= \lim_{u\to 1^-} \lim_{t\to 0^+} \frac{1 - \ln C\left(u^t, u^t\right)}{-\ln u^t} \\
&= \lim_{v\to 1^-} \frac{1 - C(v,v)}{1-v}
\end{aligned}$$

which demonstrates that C and C_e have the same coefficient of tail dependence.

The second result comes from the fact that, if the bivariate distribution has unit Frèchet marginals, by Theorem 2.3.2 there exist a finite spectral measure $\mathcal{H}$ on $\mathbb{S}^{d-1}$, such that

$$G(x_1, x_2) = \exp\left\{ - \int_{\mathbb{S}^{d-1}} \max\left(\frac{w_1}{x_1}, \frac{w_2}{x_2}\right) d\mathcal{H}(w_1, w_2) \right\},\ x_1,\ x_2 > 0,$$

with $\int_{\mathbb{S}^{d-1}} w_1 d\mathcal{H}(w_1, w_2) = \int_{\mathbb{S}^{d-1}} w_2 d\mathcal{H}(w_1, w_2) = 1$, which yields

$$\begin{aligned}
\lambda &= 2 - \int_{\mathbb{S}^{d-1}} \max(w_1, w_2)\, d\mathcal{H}(w_1, w_2) \\
&= 2 - 2\left\{ \int_{\mathbb{S}^{d-1}} \max(w_1, w_2)\, d\mathcal{H}(w_1, w_2)/2 \right\} \\
&= 2 - 2A(1/2).
\end{aligned}$$

□

PROOF. Theorem 2.5.1: we only need to prove the claim for bivariate EV copulas, since the result applies to the bivariate marginals of any higher dimensional case. By Proposition 1 in Frahm (2006) and by Theorem 7.48 in McNeil et al. (2005).

$$\begin{aligned}
\varepsilon &= \lim_{t\nearrow 1} \frac{\overline{C_e}(1-t, 1-t)}{1 - C_e(t,t)} = \lim_{t\nearrow 1} \frac{2 - (1 - C_e(t,t))/(1-t)}{2 - (2 - (1 - C_e(t,t))/(1-t))} \\
&= \lim_{t\nearrow 1} \frac{2 - \ln C_e(t,t)/\ln t}{2 - (2 - \ln C_e(t,t)/\ln t)} = \lim_{t\nearrow 1} \lim_{q\searrow 0} \frac{2 - (1 - C(t^q, t^q))/-\ln(t^q)}{2 - (2 - (1 - C(t^q, t^q))/-\ln(t^q))} \\
&= \lim_{s\nearrow 1} \frac{2 - (1 - C(s,s))/(1-s)}{2 - (2 - (1 - C(s,s))/(1-s))} \\
&= \lim_{s\nearrow 1} \frac{\overline{C}(1-s, 1-s)}{1 - C(s,s)} = \varepsilon_G
\end{aligned}$$

which shows that C and C^* share the same EDC. □

Proof. Proposition 2.5.2: by definition of ε_G

$$\begin{aligned}
\varepsilon_G &= \lim_{t\nearrow 1} \frac{\overline{C}(1-t,\ldots,1-t)}{1-C(t,\ldots,t)} \\
&= \lim_{t\nearrow 1} \frac{\sum_{K\subset D}(-1)^{|K|} C\left(t^{\mathbb{I}\{K\in I\}},\ldots,t^{\mathbb{I}\{K\in I\}}\right)}{1-C(t,\ldots,t)},
\end{aligned}$$

by equation (2.4.4) and using the asymptotic identity $\ln x \sim x-1$ we obtain

$$\begin{aligned}
\varepsilon_G &= \lim_{t\nearrow 1} \frac{\sum_{K\subset D}(-1)^{|K|} t^{\epsilon_{\mathbb{I}\{K\in I\}}}}{1-t^{\epsilon_G}}, \\
&= \frac{\sum_{K\subseteq D}(-1)^{|K|+1}\epsilon_{G_K}}{\epsilon_G}.
\end{aligned}$$

□

Proof. Proposition 2.5.3: the first inequality follows of (2.5.3) for the three dimensional case

$$\max\left\{\epsilon_{12},\epsilon_{13},\epsilon_{23},\textstyle\sum^{*}_{i,j=1,2,3}\epsilon_{ij}-3\right\} \le \epsilon_{123} \le \min_{i=1,2,3}\left\{\textstyle\sum_{k=1,2,3;k\neq i}\epsilon_{ik}-1\right\}$$

and the fact that

$$\epsilon_{ij} = \frac{2}{1+\varepsilon_{ij}} \tag{2.A.1}$$

$$\epsilon_{ijk} = \frac{3-\sum^{*}_{i,j=1,2,3}\frac{2}{1+\varepsilon_{ij}}}{\varepsilon_{ijk}-1} \tag{2.A.2}$$

by equation (2.5.1), assuming $\epsilon_i = 1$ for all $i \in K$.

The second inequality follows by lemma 6 in Schlather and Tawn (2002), which said that for an index set $K = \{i,j,k\}$ and the satisfaction of the inequality (2.5.2) then, the next inequality is valid

$$\max\{\epsilon_{ij}+\epsilon_k,\,\epsilon_{ik}+\epsilon_{ik}+\epsilon_j\} - \min\{2,\epsilon_{ij}+\epsilon_{ik}-1\} \le \epsilon_{jk} \le \min\{\epsilon_j+\epsilon_k,\epsilon_{ij}+\epsilon_{ik}-\epsilon_i\}. \tag{2.A.3}$$

Replacing equations (2.A.1) and (2.A.2) in (2.A.3) one obtains the wished result.

The third inequality is a direct result of the Theorem 3.4 in Joe (1997). Let C_{ij} be the copula of $\mathbb{P}(U_i > t,\, U_j > t)$ and $C_{i|j}$ be the copula of $\mathbb{P}(U_i > t \mid U_j > t)$ for an ease of notation, where $\mathbf{U} = (U_i, U_j.U_k)$ are uniform random variables. Then,

$$\begin{aligned}
\frac{\overline{C_{jk}}(1-t,1-t)}{1-C_{jk}(t,t)} &= \frac{C_{j|k}(t,t)}{2-C_{j|k}(t,t)} \\
&= \frac{C_{ji|k}(t,t)+C_{ji|k}(t,1-t)}{2-\left(C_{ji|k}(t,t)+C_{ji|k}(t,1-t)\right)} \\
&\le \frac{\min\left\{C_{k|i}(t,t),C_{j|i}(t,t)\right\}+1-C_{i|k}(t,t)}{2-\left(\min\left\{C_{k|i}(t,t),C_{j|i}(t,t)\right\}+1-C_{i|k}(t,t)\right)}
\end{aligned}$$

By taking limits

$$\begin{aligned}\varepsilon_{jk} &= \lim_{t\nearrow 1}\frac{\overline{C_{jk}}\left(1-t,1-t\right)}{1-C_{jk}\left(t,t\right)} \leq \lim_{t\nearrow 1}\frac{\min\left\{C_{k|i}\left(t,t\right),C_{j|i}\left(t,t\right)\right\}+1-C_{i|k}\left(t,t\right)}{2-\left(\min\left\{C_{k|i}\left(t,t\right),C_{j|i}\left(t,t\right)\right\}+1-C_{i|k}\left(t,t\right)\right)}\\ &\leq \frac{\min\left\{\lambda_{ki},\lambda_{ji}\right\}+1-\lambda_{ik}}{1-\min\left\{\lambda_{ki},\lambda_{ji}\right\}+\lambda_{ik}},\end{aligned}$$

and interchanging the subscripts i and k and from combining these two results we obtain the upper bound

$$\varepsilon_{jk} \leq \frac{1-\left|\lambda_{ki}-\lambda_{ji}\right|}{1+\left|\lambda_{ki}-\lambda_{ji}\right|}.$$

The lower bound follows similar arguments. □

PROOF. Proposition 2.5.4: by definition of the extreme value copula

$$\begin{aligned}C_H^*\left(\mathbf{t}\right)=C_H^*\left(H_1\left(x_1\right),\ldots,H_d\left(x_d\right)\right) &= C^*_{G^{\theta(\tau)}}\left(G_1\left(x_1\right)^{\theta_1},\ldots,G_d\left(x_d\right)^{\theta_d}\right)\\ &= C_G^{*\theta(\tau)}\left(G_1\left(x_1\right)^{\theta_1/\theta(\tau)},\ldots,G_d\left(x_d\right)^{\theta_d/\theta(\tau)}\right)\\ &= C_G^*\left(t_1^{\theta_1},\ldots,t_d^{\theta_d}\right),\end{aligned}$$

the equality follows. □

PROOF. Proposition 2.5.7: by definition of ε_H and using the asymptotic identity $\ln x \sim x-1$ we obtain

$$\begin{aligned}\varepsilon_H &:= \lim_{t\nearrow 1}\frac{\overline{C_H^*}\left(1-t,1-t\right)}{1-C_H^*\left(t,t\right)}\\ &= \lim_{t\nearrow 1}\frac{\overline{C_G^{*\theta(1,1)}}\left(1-t,1-t\right)}{1-C_G^{*\theta(1,1)}\left(t,t\right)}\\ &= \lim_{t\nearrow 1}\frac{1-t^{\theta_1}-t^{\theta_2}+C_G^{*\theta(1,1)}\left(t,t\right)}{1-C_G^{*\theta(1,1)}\left(t,t\right)}\\ &= \lim_{t\nearrow 1}\frac{\ln t^{\theta_1}+\ln t^{\theta_2}+\ln t^{\epsilon_H}}{\ln t^{\epsilon_H}}\\ &= \lim_{t\nearrow 1}\frac{\theta_1+\theta_2-\epsilon_H}{\epsilon_H} = \lim_{t\nearrow 1}\frac{\lambda_H}{\theta_1+\theta_2-\lambda_H},\end{aligned}$$

where $\lambda_H=\theta_1+\theta_2-\epsilon_H$ or $\lambda_H=\lim_{t\nearrow 1}\overline{C_G^{*\theta(1,1)}}\left(1-t,1-t\right)/1-t$, with $\lambda_H\in\left[0,\min\left\{\theta_1,\theta_2\right\}\right]$. □

PROOF. Proposition 2.5.10: by (2.5.5) and definition of ϵ_H we obtain for (1).

$$
\begin{aligned}
\varepsilon_H &= \lim_{t \nearrow 1} \frac{\overline{C_H^*}\left(1-t,\ldots,1-t\right)}{1-C_H^*\left(t,\ldots,t\right)} \\
&= \lim_{t \nearrow 1} \frac{\overline{C_G^{*\theta(1,\ldots,1)}}\left(1-t,\ldots,1-t\right)}{1-C_G^{*\theta(1,\ldots,1)}\left(t,\ldots,t\right)} \\
&= \lim_{t \nearrow 1} \frac{\sum_{I \subset D}\left(-1\right)^{|I|} C_G^{*\theta(1,\ldots,1)}\left(t^{\mathbb{I}_{\{1\in I\}}},\ldots,t^{\mathbb{I}_{\{d\in I\}}}\right)}{1-C_G^{*\theta(1,\ldots,1)}\left(t,\ldots,t\right)},
\end{aligned}
$$

expand the terms for the copula $C_G^{*\theta(1,\ldots,1)}\left(t\right) = \min\left(t_1^{\theta_1},\ldots,t_d^{\theta_d}\right)$ and using the asymptotic identity $\ln x \sim x-1$

$$
\begin{aligned}
\varepsilon_H &= \lim_{t \nearrow 1} \frac{\sum_{I \subset D}\left(-1\right)^{|I|} \min^{\theta(1,\ldots,1)}\left(t^{\mathbb{I}_{\{1\in I\}}},\ldots,t^{\mathbb{I}_{\{d\in I\}}}\right)}{1-\min^{\theta(1,\ldots,1)}\left(t,\ldots,t\right)}, \\
&= \lim_{t \nearrow 1} \frac{\sum_{I \subset D}\left(-1\right)^{|I|} \min\left(t^{\theta_1\mathbb{I}_{\{1\in I\}}},\ldots,t^{\theta_d\mathbb{I}_{\{d\in I\}}}\right)}{1-\min\left(t^{\theta_1},\ldots,t^{\theta_d}\right)} \\
&= \lim_{t \nearrow 1} \frac{\sum_{I \subset D}\left(-1\right)^{|I|}\left(1-\ln t^{\max_{j\in I}\theta_j}\right)}{1-\left(1-\ln t^{\max_{1\le j\le d}\theta_j}\right)} \\
&= \frac{\sum_{I \subset D}\left(-1\right)^{|I|+1}\max_{j\in I}\theta_j}{\max_{1\le j\le d}\theta_j} \\
&= \frac{\sum_{j=1}^d \theta_j - \left(\sum_{j=1}^d \theta_j - \min_{1\le j\le d}\theta_j\right)}{\max_{1\le j\le d}\theta_j}.
\end{aligned}
$$

Thus, the result for (1) follows.

The proof for (2) follows from statements (1), (3) has arguments analogous to the above proof replacing the copula $C_G^{*\theta(1,\ldots,1)}\left(t\right) = \prod_{j=1}^d\left(t_1^{\theta_1},\ldots,t_d^{\theta_d}\right)$ and (4) follows from Definition 2.5.2 and Proposition 2.5.7. □

Proof. Corollary 2.5.11: this Corollary follows immediately from Proposition 2.5.10 (1). □

CHAPTER 3

The Multivariate extremal index and the visualization of measures of extreme dependence

"Extreme positions are not succeeded by moderate ones,
but by contrary extreme positions."
(Friedrich Nietzsche)

3.1. Introduction

One of the most important assumptions underlying the currently known methods for fitting multivariate extreme value distributions is that the observations of different events are independent. However, in finance extreme events often occur simultaneously.

Such clusters of extreme events are important features in understanding risk over a given time interval. This measure can be represented by the extremal index, which can be interpreted as a measure for the frequency of extremes to cluster in the limit. In particular, a stationary sequence of random variables X is said to have an extremal index $\theta \in [0,1]$ if for every $\tau > 0$ there exists a sequence of thresholds u_n such that $n\mathbb{P}(X_1 > u_n) \to \tau$ and $\mathbb{P}\left(\bigvee_{i=1}^{n} X_i \leq u_n\right) \to \exp(-\tau\theta)$ as $n \to \infty$, with $\theta = 1$ corresponding to asymptotic independence and $\theta \to 0$ to an increasing cluster of extreme observations (Leadbetter et al., 1983). In a 1-dimensional case there are numerous techniques to calculate the extremal index and for declustering a series of observations to obtain independent extreme values. See for example Embrechts et al. (1997) and Beirlant et al. (2004), for a resume of these.

The notion of multivariate extremal index is relatively more complicated because this parameter is now a function. A first definition of this concept was made by Nandagopalan (1994). In fact, one can start from a stationary process on a d-dimensional space and reduce the strength of temporal dependence between random sequences exceeding a d-dimensional threshold sequence to the multivariate extremal index function (MEI).

A first statistical method to estimate the MEI was proposed by Smith and Weissman (1996). They proposed two methods to calculate the MEI. The first method is based on a discrete approximation of a ratio of two multivariate extreme value dependence functions. In the second method they showed that under fairly general conditions, extremal properties of a wide class of multivariate time series may be calculated by approximating the processes by a multivariate maxima of moving maxima process, and together with some mixing conditions, both processes share the same multivariate extremal index. However, the MEI is not the only functional measure of interest. For example the functional of the distribution of the excesses and the functional of the distribution of the cluster size are of interest too.

In order to apply multivariate extreme value theory some form of data filtering it is necessary to ensure independence of the marginals. A few declustering schemes have been proposed for multivariate sequences. The approach of Coles and Tawn (1991) is a multivariate version of blocks declustering by assuming that each independent event has a standard length of separation with other extreme event, while the approach of Nadarajah (2001) is a more sophisticated extension, which exploits the knowledge of the clusters size and the cluster distribution functional. In each case we need to choose one or more arbitrary parameter, which can influence drastically our estimations.

The present study makes twofold contribution. First, it puts forward a new approach to estimate extreme cluster functionals. Specifically the MEI and the cluster size probabilities. Second, we use this framework to analyse the financial crisis in South East Asia in late 1997. We found that, before and after the crisis, there was not a marked extreme dependence between the Asian markets, and among the Asian market, Japan and the U.S.A.

The layout of the chapter is as follows. Section 3.2 reviews the main concepts related to the MEI and cluster functionals. A new method for the estimation of the MEI is introduced and it does not require an arbitrary choice of declustering parameters based on a property of the time of exceedances over a sequence of thresholds. Section 3.4 discusses a new declustering method based on a prior estimation of the MEI, so that the associated declustered time series are supported by the limiting theory. Section 3.5 reviews a novel measure of tail dependence introduced by Hsing et al. (2004), which allows to visualise the extreme tail dependence and it will be of help to realize a complete analysis in the empirical section. In section 3.6 the performance of these methods are assessed through an application to real financial time series, specifically the Asian crisis. In particular, we examine price linkages among Asian equity markets in the period of crisis in 1997. Section 3.7 summarizes the conclusions and remarks of this work.

3.2. The multivariate extremal index

We need a formal definition of what we understand as multivariate extremal index. Let $\mathbf{X}$ be a d- dimensional stationary sequence with joint cumulative distribution function $F(\mathbf{x})$ belongs to the maximum domain of attraction of a multivariate extreme value distribution $H(\mathbf{x})$ ($F \in MDA(H)$). Moreover, it satisfies the long-range dependence conditions $\triangle(u_n)$ of Nandagopalan (1994) for $u_n > 0$ and $\tilde{\mathbf{X}}$ is an associated iid sequence of $\mathbf{X}$ with joint distribution function $\tilde{F} \in MDA(G)$.

Similarly, let

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{M}_n \le \mathbf{u}_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{j=1}^{d} M_{nj} \le u_{nj}\right) = \exp\{-\mu([\mathbf{0},\mathbf{x}]^c)\} = H(\mathbf{x}),$$

where $\mathbf{M}_n$ denotes the vector of the maxima of the stationary sequence $\mathbf{X}$ and

$$\lim_{n\to\infty} \mathbb{P}\left(\tilde{\mathbf{M}}_n \le \mathbf{u}_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{j=1}^{d} \tilde{M}_{nj} \le u_{nj}\right) = \exp\{-\tilde{\mu}([\mathbf{0},\mathbf{x}]^c)\} = G(\mathbf{x})$$

the vector of maxima of the associated iid sequence $\tilde{\mathbf{X}}$, where μ and $\tilde{\mu}$ are exponent measures (see Resnick (1987)). Then, the multivariate extremal index of $\mathbf{X}_n$ is defined by the relation

$$\theta\left(\tau\right)=\frac{\mu\left(\left[\mathbf{0},\mathbf{x}\right]^c\right)}{\tilde{\mu}\left(\left[\mathbf{0},\mathbf{x}\right]^c\right)}=\frac{\ln H\left(\mathbf{x}\right)}{\ln G\left(\mathbf{x}\right)}, \tag{3.2.1}$$

where $\tau=\tau\left(x_j\right)=-\ln H\left(x_j\right)$ for all $j=1,\ldots,d$.

A stationary sequence of random variables $\mathbf{X}$ is said to have a multivariate extremal index $\theta\in[0,1]$ if for every $\tau\in\mathbb{R}^d$ there exists a sequence of thresholds $\mathbf{u}_n\in\mathbb{R}^d$ such that $n\mathbb{P}\left(X_{1j}>u_{nj}\right)\to\tau_j$ for all $j=\{1,\ldots,d\}$. This result allows extending the MEV theory of iid sequences to the stationary case. A natural interpretation of the multivariate extremal index is that it describes the strength of temporal dependence in the asymptotic distribution of the vectors of componentwise maxima where $\theta=1$ corresponding to asymptotic independence and $\theta\to 0$ an increment in the frequency of large observations in cluster (Smith and Weissman (1996),Nandagopalan (1994)).

Some basic properties of the multivariate extremal index show that:

(1) $0\le\theta(\tau)\le 1$ for all $\tau>0$.
(2) $\theta\left(\tau\right)=\theta\left(c\tau\right)$ for each $\tau\in\mathbb{R}^d$ and $c>0$.
(3) $\theta_j=\lim_{\tau_i\searrow 0}\theta\left(\tau_1,\ldots,\tau_d\right)$, $\forall i\neq j$,
(4) and the bounds for a distribution with MEI $\theta\left(\tau\right)$ are

$$\Gamma\left(\tau\right)\bigvee_{j=1}^{d}\theta_j\tau_j\le\theta\left(\tau\right)\le\Gamma\left(\tau\right)\sum_{j=1}^{d}\theta_j\tau_j,$$

where $\Gamma\left(\tau\right)=\lim_{n\to\infty}n\mathbb{P}\left(\mathbf{X}_1\nleq\mathbf{u}_n\right)$, and $\tau\in\mathbb{R}^d$.

In particular, if $G\left(\mathbf{x}\right)$ and $H\left(\mathbf{x}\right)$ have independent marginals, then

$$\theta\left(\tau\right)=\sum_{j=1}^{d}\theta_j\tau_j/\sum_{j=1}^{d}\tau_j. \tag{3.2.2}$$

On the other hand, if $G\left(\mathbf{x}\right)$ and $H\left(\mathbf{x}\right)$ have completely dependent marginals, then

$$\theta\left(\tau\right)=\bigvee_{j=1}^{d}\theta_j\tau_j/\bigvee_{j=1}^{d}\tau_j. \tag{3.2.3}$$

These two results enable us to have a first approximation to the dependence between random variables of the stationary sequence $H\left(\mathbf{x}\right)$ and the associated iid sequence $G\left(\mathbf{x}\right)$.

The key point of the estimation of the MEI is to express this as a ratio of two multivariate Pickands dependence functions. The next definition shows that the multivariate extremal index of a stationary sequence for a fixed τ can be estimated by univariate methods. Thus, under some conditions of accuracy approximation, we can have a discrete approximation of the function $\theta(\tau)$.

Definition 3.2.1. (Smith and Weissman (1996)). Let $\mathbf{X}$ be a d-dimensional stationary sequence with unit Frèchet marginals, i.e., $F_j(x)=e^{-1/x}$ for $\mathbf{x}\in\mathbb{R}^d$, and multivariate extremal index $\theta(\tau)$. Define the univariate stationary sequence $\chi_i(\tau)=\max_{i\geq 1}\tau_d X_{id}$. Then $\theta(\tau)$ is the extremal index of the sequence $\{\chi_i(\tau)\}_{i\geq 1}^{i=n}$.

This definition shows that for a fixed τ we can obtain the 1-dimensional extremal index of the sequence $\{\chi_i(\tau)\}$, and therefore, it can be estimated by univariate methods as the interval estimator proposed by Ferro and Segers (2003).

Before we give a formal definition of this idea, let us replace the threshold of exceedances by measurable sets

$$S_n : \;=\; \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \nleq \mathbf{u}_n\right\} \tag{3.2.4}$$

of extreme observations in at least a marginal, that is, $X_{i,j} > x_j$ for some $j = 1, \ldots, d$ and $T_{S,n} := \mathbb{P}\left(\mathbf{X_i} \notin S_n \mid \mathbf{X_1} \in S_n\right)$ for $2 \leq i \leq n$, the time between inter-exceedances. Moreover, let $p_{S_n} = \mathbb{P}\left(\mathbf{X_1} \in S_n\right)$.

Let us define the subsets

$$W_n = (-\infty, u_n]^c, \tag{3.2.5}$$

and the triangular arrays $B_{1,n}, \ldots, B_{r,n}$. Then, $B_{i,n}(\tau) = \{\chi_i(\tau) \in W_n\}$ is defined with each row composited of block-stationary events, that is, $\mathbb{P}\left(\bigcup_{i=1}^m B_{i+k,n}(\tau)\right) = \mathbb{P}\left(\bigcup_{i=1}^m B_{i,n}(\tau)\right)$ for every $m = 1, \ldots, r-1$, $k = 1, \ldots, r-m$ and $n \geq 1$. This condition is relatively weaker than the assumption that the vectors of indicator variables $I_i = I\left(B_i(\tau)\right)$ are strictly stationary.

For notational convenience, all right superscripts are used as indexes (not powers) throughout the rest of this section. Denote $p^\tau_{m,n} = \mathbb{P}\left(\bigcup_{i=1}^m B_{i+k,n}(\tau)\right)$ the probability of non observe an extreme event in a block of length m, $q^\tau_{m,n} = 1 - p^\tau_{m,n}$, $p^\tau_n = p^\tau_{1,n}$ and the mixing condition

$$\begin{aligned}\alpha^\tau_{s,l,n} \;=\; \max\Bigg\{&\left|\mathbb{P}\left(\bigcap_{i=u+1}^{v} B_i^c(\tau) \cap \bigcap_{k=s+v+1}^{s+w} B_k^c(\tau)\right) - q^\tau_{v-u} q^\tau_{w-v}\right| \\ &: u \geq 0,\, v-u \geq l,\, w-v \geq l,\, w+s \leq r \Bigg\},\end{aligned}$$

representing the strength of the dependence between arrays.

Let $T_{B,n}(\tau) = \min\{i = 1, \ldots, r_n : \mathbb{I}_{i+1,n} = 1 \mid \mathbb{I}_{1,n} = 1\}$ be a random variable representing the times between exceedances in $\chi_n(\tau)$ for a fixed $\tau > 0$ with distribution

$$\mathbb{P}\left(T_{B,n}(\tau) \geq t \mid B_1(\tau)\right) = \mathbb{P}\left(\bigcap_{i=1}^{t-1} B^{\tau c}_{i+1} \mid B_1\right) = \theta_{t,n}(\tau), \tag{3.2.6}$$

where $\mathbb{I}_{i,n}$ is the indicator of the event $B_{i,n}(\tau)$ for $n \geq 1$. A simple interpretation of (3.2.6) is as the probability of that the time between two extreme events will be at least $t-1$. This result allows extending the one dimensional interval estimator of Ferro and Segers (2003) to the multidimensional case in the next theorem.

THEOREM 3.2.2. *Let $\mathbf{X}$ be a* d*-dimensional stationary sequence with unit Frèchet margins, i.e., $F_j(x) = e^{-1/x}$, for $\mathbf{x} \in \mathbb{R}^d$ and multivariate extremal index $\theta(\tau)$. Define the univariate stationary sequence $\chi_i(\tau) = \max_{i\geq 1}\tau_d X_{id}$, with $\sum_{j=1}^d \tau_j = 1$ and $\tau_j \in [0,1]$ for all $j = 1, \ldots, d$. Then $\theta(\tau)$ is the extremal index of the sequence $\{\chi_i(\tau)\}_{i\geq 1}^{i=n}$. Hence, the normalized inter-arrival time between extreme events in $\mathbf{X}$ is approximately distributed as follows*

$$\mathbb{P}\left(p_{S_n} T_{S,n} > x\right) \to \mathbb{P}\left(p^\tau_n T_{B,n}(\tau) \geq x\right) = \theta_n(\tau) \exp\left(-x\theta_n(\tau)\right) + o(1), \tag{3.2.7}$$

for $x > 0$.

The multivariate extremal index $\theta(\tau)$ is one of many other measures of extreme dependence, exploring other characteristics we can obtain other interesting properties of the stationary sequence $\mathbf{X}$, such as the cluster size, the peak of the excesses, or the sum of all excesses.

Yun (2000) called this type of measures cluster functionals and he showed how to find some limit distributions for the high order Markow chain case. A more general extension to stationary sequences was described by Segers (2003). Now, we will study some statistics based on such cluster functionals.

Definition 3.2.3. (Yun (2000)) A measurable map $c : \mathbb{R} \cup \mathbb{R}^2 \cup \mathbb{R}^3 \cup \cdots \to \mathbb{R}$ is called a cluster functional if for all integers $1 \leq u \leq v \leq r$ and for all $(x_1, \ldots, x_r)$ such that $x_i \leq 0$ whenever $i = 1, \ldots, u-1$ or $i = k+1, \ldots, r$ we have $c(x_1, \ldots, x_r) = c(x_u, \ldots x_v)$.

A natural definition of cluster functionals can be written in terms of the distribution of the stationary sequence conditional on the event that the first variable belongs to a subset like W_n. Roughly speaking, these are functionals which depend on all shortest vectors of observations containing exceedances over a given high threshold. An asymptotic theory on estimators of functionals of that type would be a significant step forward towards a general approach to analyze the extremal dependence structure of stationary time series.

Definition 3.2.4. (Segers (2003)) Let $\{\chi_i(\tau)\}$ be stationary for a $\tau \in [0,1]^d$. For integers $1 \leq u \leq v \leq r$ as in Definition 3.2.3 such that the next condition is accomplished

$$\lim_{m\to\infty} \limsup_{n\to\infty} \sum_{i=m+1}^{r} \mathbb{P}(\chi_i(\tau) > u_n \mid \chi_1(\tau) > u_n) = 0$$

so that the expected cluster size is uniformly bounded. Then, for every sequence of cluster functionals c_n and measurable sets $A_n \subset \mathbb{R}$ we have

$$\begin{aligned} &\mathbb{P}\left(c_n(\chi_i(\tau) - u_n)_{i=1}^r \in A_n \mid \textstyle\bigvee_{i=1}^r \chi_i(\tau) > u_n\right) \\ &\sim \theta(\tau)^{-1}\left[\mathbb{P}\left(c_n(\chi_i(\tau) - u_n)_{i=1}^t \in A_n \mid \chi_1(\tau) > u_n\right)\right. \\ &\left.-\mathbb{P}\left(c_n(\chi_i(\tau) - u_n)_{i=2}^t \in A_n, \textstyle\bigvee_{i=1}^t \chi_i(\tau) > u_n \mid \chi_1(\tau) > u_n\right)\right] \end{aligned} \tag{3.2.8}$$

as $n \to \infty$, where $\theta(\tau)$ is the multivariate extremal index for a fixed $\tau > 0$.

The most of cluster functionals of our interest have the form

$$c_n(x_1, \ldots, x_r) = \sum_{i=-m+2}^{r} \phi_n(x_i, \ldots, x_{i+m-1}), \tag{3.2.9}$$

for all $i = 1, \ldots, m$ where $\phi_n : \mathbb{R}^m \to [0, \infty)$ is a measurable function such that $\phi_n(x_1, \ldots, x_m = 0)$ if $\bigvee_{i=1}^m x_i \leq 0$.

Cluster functionals based on this representation (3.2.9) are for example for $m = 1$ and $\phi_n(\chi_i(\tau)) = \mathbb{I}\{B_{i,n}(\tau)\}$ the number of exceedances, for $m = 1$ and $\phi_n(\chi_i(\tau)) = \max(x, 0)$ we obtain the sum of all excesses.

A cluster functional that is not of this type is the duration of a cluster that corresponds to the functional

$$c_n(x_1, \ldots, x_r) = \max\{j - i + 1 : 1 \leq i \leq j \leq r,\, x_i > 0,\, x_j > 0\}.$$

Moreover, following Leadbetter et al. (1983) for the one dimensional case we define the functional of the cluster size distribution for the multivariate case in function of the associated stationary sequence $\{\chi_i(\tau)\}$ in the next definition.

DEFINITION 3.2.5. Let $\mathbf{X}$ be a d-dimensional stationary sequence with unit Frèchet margins and MEI $\theta(\tau)$, and $\chi_i(\tau) = \max_{i\geq 1}\tau_d X_{id}$ the associated stationary sequence for $\tau \in [0,1]^d$ with multivariate extremal index $\theta(\tau)$. Then the functional of cluster size distribution of $\mathbf{X}$ can be written in terms of $\chi_i(\tau)$ as follows

$$\begin{aligned}\pi_n(\tau,k) &= \mathbb{P}\left[\sum_{i=1}^{r_n}\mathbb{I}\{B_{i,n}(\tau)\}\geq k \mid \bigvee_{i=1}^{r_n}\chi_i(\tau)\in W_n\right]\\ &= \theta(\tau)\,\mathbb{P}\left[\sum_{i=1}^{r_n}\mathbb{I}\{B_{i,n}(\tau)\} = k-1 \mid \chi_1(\tau)\in W_n\right]+o(1)\end{aligned}$$

where W_n is the set of extreme events and $r_n = o(n)$.

Furthermore, Hsing et al. (1988) give extra assumptions for a fixed τ under which we obtained the mean of the limiting cluster size distribution $\theta(\tau)^{-1} = \lim_{n\to\infty}\sum_{k=1}^{\infty} k\pi_n(\tau,k)$.

3.3. Estimators of multivariate extremal index

Unlike the interval estimator proposed by Ferro and Segers (2003), here we exploit the fact that two consecutive exceedances follow a point mass-exponential mixture distribution. If we display $p_n^\tau T_{B,n}(\tau)$ against standard exponential quantiles, we will obtain one segment composed of zeros and a line with gradient $\theta(\tau)^{-1}$, which intersect each other at $(-\log\theta(\tau),0)$. To ensure that the contributions from the smallest inter-exceedance times are indeed zero we use $p_n^\tau(T_{B,n}(\tau)-1)$ instead of $p_n^\tau T_{B,n}(\tau)$.

We normalize the collection of inter-exceedance times by its means, and order them in increasing order, which define a new sequence$\{Z_i\}_{i=1}^{N-1}$. Thus, we can compare this with a collection of standard exponential quantiles defined as $q_i = -\log(1-i/(N-1))$.

If the sequence $\{Zi, q_i\}$ corresponds to the limiting distribution of $p_n^\tau(T_{B,n}(\tau)-1)$ we can hope that the resultant model is of the form $Z = \left(q_i - \log\theta(\tau)^{-1}\right)/\theta(\tau)$. Applying a weighted least square procedure to the next model

$$f\left(\theta^{-1}\right) = \left\{\sum_{i=1}^{N-1} p_i\left(\max\left(Z_i - \left(q_i - \log\theta^{-1}\right)/\theta, 0\right)\right)^2\right\}, \tag{3.3.1}$$

where $p_i = \left(\sum_{j=N-i}^{N-1} j^{-2}\right)^{-1}$ are the weights choosing of such form, so that Z_i is an ordered sample. Thus, the weighted least square estimator for the inter-exceedance times is defined as

$$\tilde{\theta}_n^{wls} = \arg\min\left\{f\left(\theta^{-1}\right)\right\}^{-1}. \tag{3.3.2}$$

Note that using the exponential structure of the inter-exceedances times and the fact that a proportion of $(1-\theta)$ exceedances constitute other cluster, then, choosing from the original sequence $\{Z_i\}$ the largest $\left\lfloor\tilde{\theta}_n^{wls}(N-1)\right\rfloor$ gap produces a new set of inter-exceedances times obtained from the first estimation. Thus, an iterative method using equations (3.3.1) and (3.3.2) can be used until it reaches a limit, where there are no differences between the gaps.

Just as the limiting distribution of the inter-exceedance times has been used to derive the interval estimator for the extremal index in Ferro and Segers (2003), so the limiting distribution if $\{\chi_i(\tau)\}$ is m-dependent can be used to derive an estimator for the cluster size distribution.

$$\pi_n(\tau,k) = e_k - \sum_{i=1}^{k-1} e_{k-i}\pi_n(\tau,i),$$

for $j \geq 2$ and $\pi_n(\tau,1) = e_1$, where an estimate for e_k based on the time of exceedances is

$$\tilde{e}_k = \frac{2(N-1)\sum_{i=1}^{N-1-k} \min(T_{B,i}-1,\ T_{B,i+k}-1)}{(N-1-k)\sum_{i=1}^{N-1}(T_{B,i}-1)}.$$

Consistency of the interval estimator for m-depended processes can be found in Ferro and Segers (2003).

Example 3.3.1. Consider a Multivariate maxima of moving maxima process (M_4), $\mathbf{X}_n = X_{n,j} = \bigvee_{l\geq 1}\bigvee_{\infty<k<-\infty} \alpha_{lkj}\psi_{l,n-l}$, for $j = 1,\ldots,3$, $l = 1$ and $1 \leq k \leq 3$, where $\{\alpha_{111},\ldots,\alpha_{131}\} = \frac{1}{3}$, $\{\alpha_{111},\alpha_{121}\} = \frac{1}{2}$ and $\alpha_{113} = 1$ are non negative constants satisfying $\sum_{l\geq 1}\sum_{\infty<k<-\infty}\alpha_{lkj} = 1$ for $j = 1,\ldots 3$ and ψ_{ln} is a vector of iid unit Frèchet random variables. The M_4 process was introduced by Smith and Weissman (1996) and it has the particularity that we can derive the multivariate extremal index easily. The extremal index of each marginal is clearly $\theta_1 = 1$, $\theta_2 = \frac{1}{2}$ and $\theta_3 = \frac{1}{3}$, hence the cluster size are $1, 2$ or 3.

For ease of the illustration we calculate first the bivariate extremal index of $\mathbf{X}_n = \{X_{n,1}, X_{n,2}\}$. Following Smith and Weissman (1996) we calculate the bivariate extremal index by means of Pickand's dependence function. Observe that doing $\frac{1}{u} = \frac{1}{x_1} + \frac{1}{x_2}$ we get

$$\begin{aligned}
F(x_1,x_2) &= \mathbb{P}(X_{i,1} \leq x_1,\ X_{i,2} \leq x_2) \\
&= \mathbb{P}(\psi_i \leq 2x_2 \wedge 3x_1,\ \psi_{i+1} \leq 2x_2 \wedge 3x_1,\ \psi_{i+2} \leq 3x_1) \\
&= \exp\left\{-\frac{1}{u}\left[\frac{(2w+1)}{3} \vee (1-w)\right]\right\}
\end{aligned}$$

and for $u \to \infty$

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{P}(M_{1,n} \leq u) &= \lim_{n\to\infty}\mathbb{P}\left(\bigvee_{i=1}^{n} X_{i,1} \leq u, \bigvee_{i=1}^{n} X_{i,2} \leq u\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\bigvee_{i=1}^{n} \psi_i \leq \frac{2u}{w} \wedge \frac{3u}{(1-w)}\right) \\
&\quad \mathbb{P}\left(\psi_{n+1} \leq \frac{2u}{w} \wedge \frac{3u}{(1-w)}\right)\mathbb{P}\left(\psi_{n+2} \leq \frac{3u}{(1-w)}\right) \\
&\to \exp\left\{-\left[\frac{2}{w} \vee \frac{(1-w)}{3}\right]\right\}
\end{aligned}$$

the limit follows.

From the definition of the extremal index we know that $\mathbb{P}(M_n \le u) \approx F^{\theta(w)}(u)$, when $u_n \to \infty$, so that

$$\begin{aligned}\theta(w) &\approx \frac{\log \mathbb{P}(M_n \le u)}{n \log F(u)} \\ &= \begin{cases} \frac{1}{3} & 0 < w \le \frac{2}{5} \\ \frac{3w}{2(2w+1)} & 0 < w \le 1. \end{cases}\end{aligned}$$

Moreover, it is easy to prove, that the limiting cluster size probabilities for the sizes 2 and 3 satisfies

$$\pi(2,w) = \begin{cases} 0 & 0 < w \le \frac{2}{5} \\ \frac{5w-2}{2w+1} & 0 < w \le \frac{2}{5} \end{cases}, \text{ and } \pi(3,w) = \begin{cases} 1 & 0 < w \le \frac{2}{5} \\ \frac{3(1-w)}{2w+1} & 0 < w \le \frac{2}{5}. \end{cases}$$

To show the quality of our estimator we simulate 10.000 realizations of this bivariate process $(X_{n,1}, X_{n,2})$. The Figure 3.3.1 displays the results of the simulation. The bivariate extremal index is calculated as in (3.3.2) for $w_i = i/40$, $i = 1, \ldots, 40$ and smoothed by averaging this with a windows sample of size $k = 3$, i.e.

$$\hat{\theta}_{n,k}(w_i) = \frac{1}{2k+1} \sum_{j=-k}^{k} \hat{\theta}_n(w_{i-j}).$$

This result is displayed on the left-top of the figure. The confidence intervals are calculated as in Ferro and Segers (2003). The grey line is the true extremal index, while the black line is the estimated. The dashed lines display the confidence interval.

The plot on the right-top depicts the root of the mean squared error (MSE)

$$MSE(w_i) = \frac{1}{m} \sum_{j=1}^{m} \left(\hat{\theta}^{(j)}_{n,k}(w_i) - \theta(w_i)\right)^2,$$

where $m = 200$ is the number of estimations of $\hat{\theta}_{n,k}(w_i)$ based on m different simulations of the bivariate process $(X_{n,2}, X_{n,3})$ of sample size 10.000.

The bottom-left plot displays the estimations of the cluster size probabilities of size two and three, and the bottom-right plot is the estimated root of the MSE. As we can observe the results are very encouraging.

EXAMPLE 3.3.2. Let us now extend this example to the trivariate case $\mathbf{X}_n = \{X_{n,1}, X_{n,2}, X_{n,3}\}$. For $\frac{1}{u} = \sum_{j=1}^{3} \frac{1}{x_j}$ and $w_k = \frac{u}{x_i x_j}$ for $k \ne \{i,j\}$ we have that

$$\begin{aligned}F(x_1,x_2,x_3) &= \mathbb{P}(X_{i,1} \le x_1,\ X_{i,2} \le x_2,\ X_{i,3} \le x_3) \\ &= \mathbb{P}(\psi_i \le 1 \wedge 2x_2 \wedge 3x_1,\ \psi_{i+1} \le 2x_2 \wedge 3x_1,\ \psi_{i+2} \le 3x_1) \\ &= \exp\left\{-\frac{1}{u}\left[\frac{w_1}{3} \vee \frac{w_2}{2} \vee w_3 + \frac{w_1}{3} \vee \frac{w_2}{2} + \frac{w_1}{3}\right]\right\}\end{aligned}$$

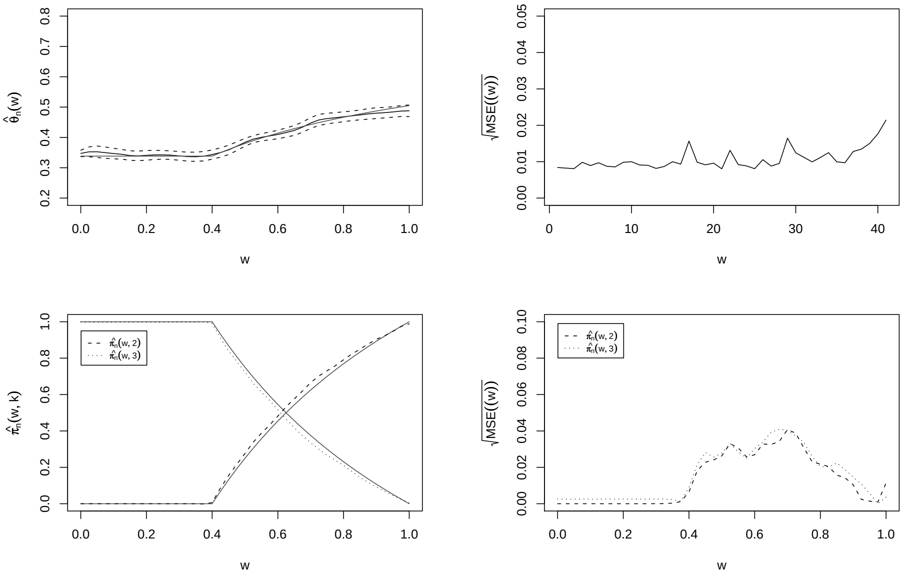

FIGURE 3.3.1. Results of the bivariate extremal index (top-left) and cluster size probabilities (bottom-left) estimations for the Example 3.3.1.

and

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(M_{1,n} \le u\right) &= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{j=1}^{3}\bigvee_{i=1}^{n} X_{i,j} \le u,\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigvee_{i=1}^{n} \psi_i \le \frac{3u}{w_1} \vee \frac{2u}{w_2} \vee \frac{u}{w_3}\right) \\
&\quad \mathbb{P}\left(\psi_{n+1} \le \frac{3u}{w_1} \vee \frac{2u}{w_2}\right)\mathbb{P}\left(\psi_{n+2} \le \frac{3u}{w_1}\right) \\
&\to \exp\left\{-\left[\frac{w_1}{3} \vee \frac{w_2}{2} \vee w_3\right]\right\}
\end{aligned}$$

and of this form we get the estimation of the trivariate extremal index

$$\theta\left(\mathbf{w}\right) = \theta\left(w_1, w_2, w_3\right) = \frac{\frac{w_1}{3} \vee \frac{w_2}{2} \vee w_3}{\frac{w_1}{3} \vee \frac{w_2}{2} \vee w_3 + \frac{w_1}{3} \vee \frac{w_2}{2} + \frac{w_1}{3}}, \tag{3.3.3}$$

where $w_j \in [0,1]$ for $j = 1, \ldots, 3$ and $\sum_{j=1}^{3} w_j = 1$.

Figure 3.3.2 displays a simulation for the last example with 10.000 realizations of this trivariate process. The differences between the true and the estimated MEI are almost imperceptible. The maximum MSE estimated is the order of 0.001 for this example.

REMARK 3.3.3. Until now we have estimated the multivariate extremal index in a set of the form $S_d = \left\{\mathbf{w} \ge 0 : \sum_{j=1}^{d} w_j\right\}$. However, this form is not restrictive. Other alternative is to estimate this in polar coordinates using the property $\theta\left(\mathbf{w}\right) = \theta\left(c\mathbf{w}\right)$ for $c > 0$. Setting

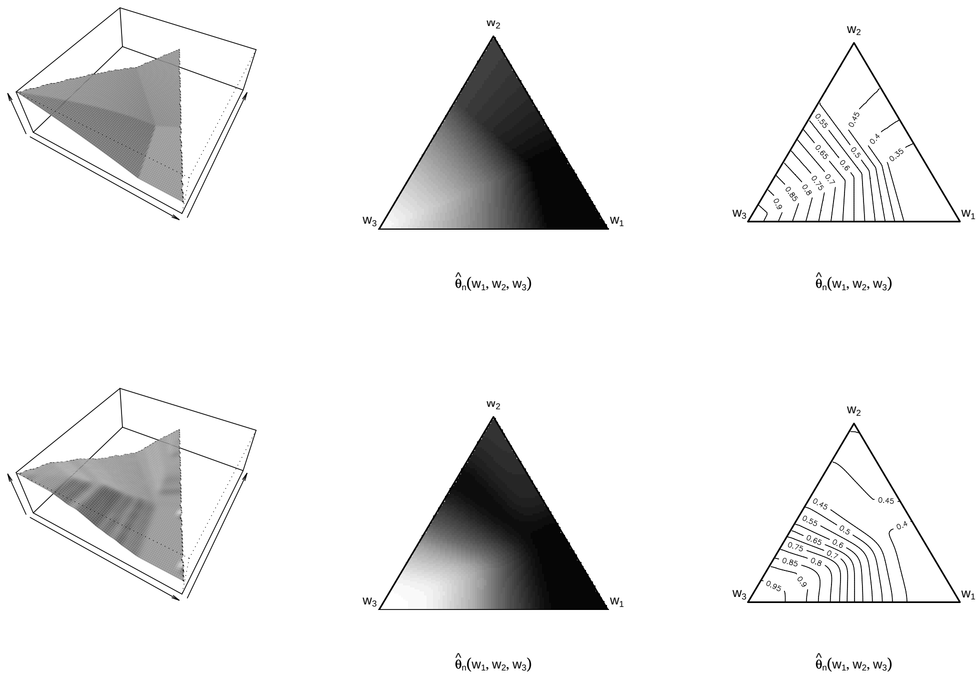

FIGURE 3.3.2. Top panel: The true trivariate extremal index of a M_4 process in Example 3.3.2. Bottom panel: estimation of the trivariate extremal index.

$c = w_1^{-1}$ as follows

$$\begin{aligned}\theta(\mathbf{w}) &= \theta(w_1,\ldots,w_d) = \theta\left(1, {}^{w_2}\!/_{w_1}\ldots, {}^{w_d}\!/_{w_1}\right)\\ &= \theta(1,\tan\varphi_1,\ldots,\tan\varphi_{d-1})\\ &= \theta(\tan\varphi_1,\ldots,\tan\varphi_{d-1}),\ \text{for } \varphi_j\in[0,\pi/2]\end{aligned}$$

with the convention $0/0 = 0$ and $\infty/0 = \infty$.

EXAMPLE 3.3.4. We consider the M_4 process in the Example 3.3.1. The trivariate extremal index for this process in polar coordinates is derived easily of (3.3.3) as follows

$$\theta(\varphi) = \theta(\varphi_1,\varphi_2) = \frac{\frac{1}{3}\vee\frac{\tan\varphi_1}{2}\vee\tan\varphi_2}{\frac{1}{3}\vee\frac{\tan\varphi_1}{2}\vee\tan\varphi_2+\frac{1}{3}\vee\frac{\tan\varphi_1}{2}+\frac{1}{3}}, \text{for } \varphi_j\in[0,\pi/2]\,.$$

Figure 3.3.3 shows the simulation for this estimator. The use of the estimator with polar coordinate will be more clear, when we introduce a tail dependence function in the section 3.5.

3.4. The multivariate declustering procedure

Multivariate declustering techniques play a very important role in extreme value theory for stationary sequence. Methods for one dimensional case are well established (see Coles (2001); Laurini and Tawn (2003) for more references). However, there are much less studies on the multivariate case. Only two procedures are known for the author; Coles and Tawn (1991) and Nadarajah (2001). The first approximation is a multivariate version of the blocks method;

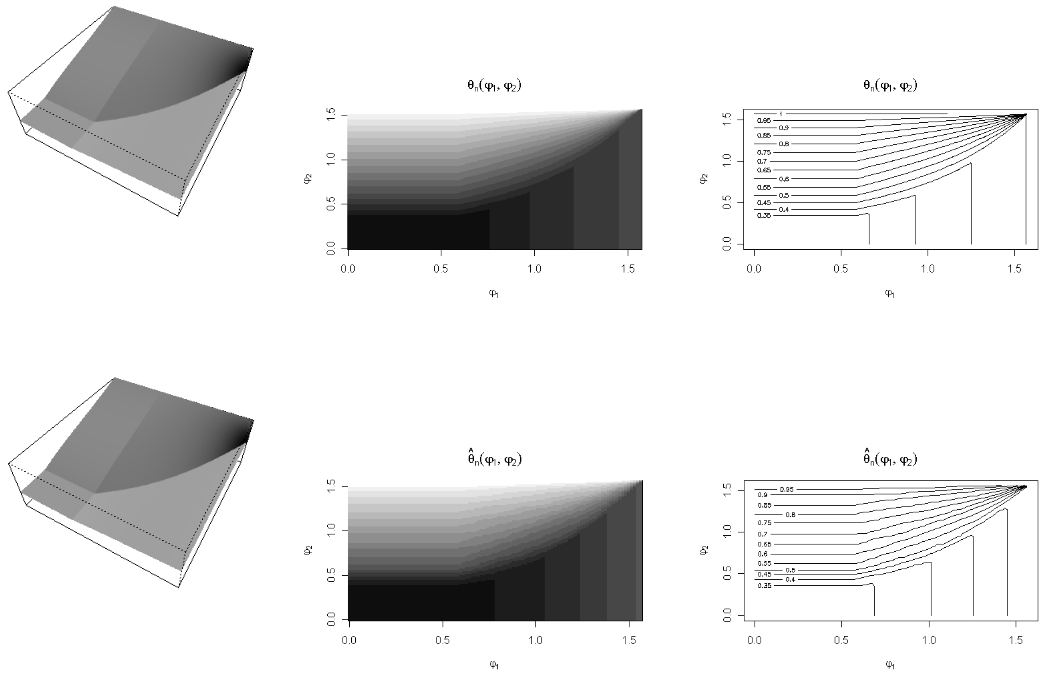

FIGURE 3.3.3. Trivariate extremal index of the M_4 process in Example 3.3.1 in polar coordinates.

independent events are filtered out by concatenating the maxima of the processes within each block. The second approximation can be considered an extension of the run method. The two methods however need the choice of one or more declustering parameters. The method proposed below uses the information provided by the multivariate extremal index to perform this task.

The difficulty in declustering a multivariate stationary sequence is due to the fact that MEI is now a function which is influenced by the dependency among marginals, that is, the extreme value copula. Following Ferro and Segers (2003) the limiting distribution (3.2.7) may be used to identify clusters without making an arbitrary choice.

If $\theta_n(\tau)$ is an estimate of the extremal index for fixed $\tau \in [0,1]^d$, then the largest $n_c - 1 = \lfloor (N-1)\theta_n(\tau) \rfloor$ of the interexceedance times $T_{i,B}$, $1 \le i \le N-1$, are approximately intercluster times. This defines a partition of the remaining interexceedance times into sets of intracluster times. By Theorem 2.2.22, the point process of exceedance times is compound Poisson, the intercluster times are independent of one another, and the sets of intracluster times are independent both from another and from the intracluster times.

Concretely, if T_{n_c} is the n_c-th largest interexceedance time and $T_{i_j,B}$ is the j-th interexceedance time to exceed $T_{n_c,B}$. Then, $\left\{T_{i_j,B}\right\}_{j=1}^{n_c-1}$ is a set of approximately independent intercluster times. In the case of ties, decrease n_c until $T_{n_c-1,B} > T_{n_c,B}$. Let also $\mathcal{T}_j = \left\{T_{i_{j-1}+1}, \ldots, T_{i_j-1}\right\}$, where $i_0 = 0$, $i_{n_c} = N$ and $\mathcal{T}_j = \emptyset$ if $i_j = i_j + 1$. Then, $\{\mathcal{T}_j\}_{j=1}^{n_c}$ is a collection of approximately independent sets of intracluster times. If we estimate θ with θ_n^{inter}, then this approach declusters the data into n_c clusters without requiring an arbitrary selection of auxiliary parameter and it is justified by the limiting theory.

The multivariate decluster procedure for a d-variate stationary sequence to obtain independent extreme events is as follows.

Assume that we can estimate for each univariate marginal about a threshold $\mathbf{u}_n = \{u_{n1}, \ldots, u_{nd}\}$ the extremal index and the independent extreme events $\tilde{X}_{n_j j}$ for $j = 1, \ldots, d$ as in Ferro and Segers (2003). Furthermore, let S_i be the time of at least one extreme event in the multivariate sequence X_j for $j = 1, \ldots, d$. Calculate the multivariate extremal index for the sequence $\chi_n(\tau)$ with $\tau = \{1/d, \ldots 1/d\}$. Applying the procedure of Ferro and Segers (2003) to identify approximately independent clusters in the sequence $\chi_n(\tau)$ we obtain the number of clusters n_χ, the functional of the cluster size, the excess over threshold and the cluster maxima. Note, that we can estimate the associated iid sequence the extreme events $\tilde{\chi}_n(\tau)$, however, the choice of the independent events $\tilde{\chi}_n(\tau)$ could not coincide with the independent extreme events in the marginals $\tilde{X}_{n_j j}$ of the stationary sequence, because of the differences in the magnitude of the marginals.

Hence, the idea is to maximize the number of independent extreme events in each dimension and in the multivariate stationary sequence, and at the same time the dependence between the marginals.

Now, with the help of the estimated multivariate extremal index $\theta_n(\tau)$ and the number of independent clusters n_χ, we can decluster the sequence S_i by identifying the cluster maxima corresponding to the most extreme value of one of the marginals or of the variable of interest.

For illustration of this procedure we observe the stationary sequence in Figure 3.4.1. The firsts three graphics show the extreme events identified in each margins by the interval decluster method to produce independent events. Let us define these independent events as $Y_{j,t}$, where j indicates the j- marginal and t the time of occurrence of the peaks. Thus, for example $Y_{1,4}$ corresponds to the first cluster maxima event in the first margin at the time 4.

The fourth graphic displays the extreme events identified in the 1-dimensional process $\chi_n(\tau)$ with $\tau = \{1/3, \ldots 1/3\}$. In the bottom panel we show the process $M_n(\mathbf{u}_n) = \bigvee_{j=1}^{3} X_{n,j}$. Observe, that the clusters in different marginals are strongly dependent in time for the first three graphics, so that the cluster maxima $Y_{j,t}$ cannot be assumed independent between marginals.

Through the Smith's Theorem 3.2.1 and Theorem 3.2.2 we can use the sequence $\chi_n(\tau)$ to approximate $\theta_n(\tau)$ for a fixed $\tau = \{1/3, \ldots 1/3\}$. With the estimator of $\theta_n(\tau)$ an one dimensional decluster procedure follows immediately for $\chi_n(\tau)$ to obtain $Y_{\chi_n,t}$ and n_c cluster without requiring an auxiliary parameter.

With this method we form a series of d-variate pseudo sequence of independent extreme events represented in the last graphic by y. However, it is possible that events, which are independent in one marginal are not taken into account, because of the difference of magnitudes in the sequence $\chi_n(\tau)$, as are the events $\{Y_{1,36}, Y_{1,92}\}$.

To include these events in the multivariate decluster procedure we observe if the times of this cluster maxima are independent from the cluster maxima obtained in the fourth graphic. This can be inferred if there are not other extreme events in the interval $[t-r, t+r]$ for the cluster maxima t in the marginal k for all $k \neq j : \{1, \ldots, d\}$. For this example $r = 2$. Thus, these events can be included in the multivariate sequence of the pseudo iid random variables.

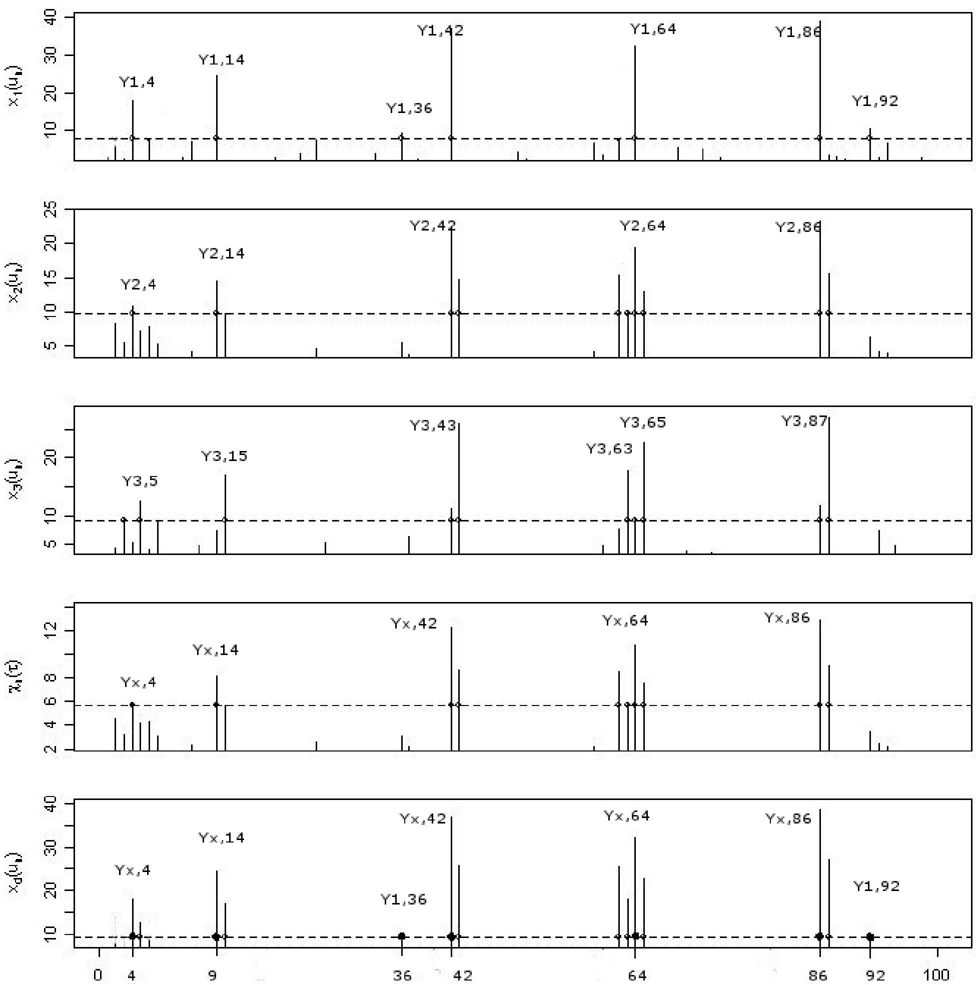

FIGURE 3.4.1. An example of a multivariate decluster procedure.

Notice, that the events $Y_{x,t}$ are a concatenation of a d-variate sequence, for example for $Yx_{,64} = \{Y_{1,64}, Y_{2,64}, X_{3,64}\}$. An important characteristic and consequence of the decluster method is that for the marginal $j = 3$, the maximum of each cluster does not coincide with the cluster maxima of the multivariate sequence and it cannot be also considered as independent. Naturally this example shows an extreme case, but it is not impossible.

For this case we propose two solutions as it is made in Nadarajah (2001). The first is replacing the sequence $X_{n,3}$ for a equivalent lagged-sequence $X'_{n,3} = X_{n-1,3}$. Thus, the cluster maxima will coincide with cluster maxima of the multivariate sequence. The second solution is to take simply the cluster maxima of the sequence $X_{n,3}$ and to conserve the cluster maxima of the multivariate sequence for the other marginals. However, this procedure can include events which are not actually realized, as in the block method (see Coles and Tawn (1991)).

3.5. The new measure of tail dependence

Other important problems in multivariate extreme value theory are the dependence between d-variate sequences. An increasing notion for modelling dependences in multivariate analysis is the concept of *Copula.*

Let $\mathbf{X} = \{X_{n1}, \ldots, X_{nd}\}$ be an iid d-variate random vector with common distribution function $F(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^d$ with continuous margins. Suppose that there exist norming constants $a_{nj}$and $b_{nj} \in \mathbb{R}^+$ for $j = 1, \ldots, d$, such that the sequence of distribution functions

$$\mathbb{P}\left(\bigcap_{j=1}^{d} M_{nj} \le x_j b_{nj} + a_{nj}\right),$$

where $M_{ij} = \bigvee_{i=1}^{n} X_{ij}$, converges to a limit distribution $G(\mathbf{x})$ with non-degenerate margins, i.e.

$$\lim_{n\to\infty} F^n\left(x_j b_{nj} + a_{nj}\right) = G(\mathbf{x}) \tag{3.5.1}$$

for all $\mathbf{x} \in \mathbb{R}^d$. In addition, all univariate marginal distribution functions are also extreme value distributions

$$G(\infty, \ldots, x_j, \ldots, \infty) = \exp\left\{-\left(1 + \gamma_j x_j\right)^{-1/\gamma_j}\right\}$$

for some $\gamma_j \in \mathbb{R}$.

Note, that the limit distribution (3.5.1) can be written as

$$\lim_{n\to\infty} n\left(1 - F\left(x_j b_{nj} + a_{nj}\right)\right) = -\ln G(\mathbf{x}).$$

Since that $\mathbb{P}(\cdot)$ is a monotone function, a continuous version of the limit remains the same.

$$\lim_{t\to\infty} t\left(1 - F\left(x_j b_j(t) + a_j(t)\right)\right) = -\ln G(\mathbf{x}). \tag{3.5.2}$$

Einmahl et al. (2001) proposes for the bivariate case a polar coordinate representation such as follows

$$G\left(\frac{x_1^{\gamma_1} - 1}{\gamma_1}, \frac{x_2^{\gamma_2} - 1}{\gamma_2}\right) = \exp\left\{-\int_0^{\pi/2}\left(\frac{1 \wedge \tan\varphi}{x_1} \vee \frac{1 \wedge \cot\varphi}{x_2}\right)\Phi(d\varphi)\right\}$$

where Φ is a finite measure on $[0, \pi/2]$, with the condition that

$$\int_0^{\pi/2}(1 \wedge \tan\varphi)\,\Phi(d\varphi) = \int_0^{\pi/2}(1 \wedge \cot\varphi)\,\Phi(d\varphi) = 1.$$

The limit (3.5.2) becomes simpler if we adopt the representation of Einmahl et al. (2001), which yields to

$$\begin{aligned}
&\lim_{t\to\infty} t\mathbb{P}\left(1 - F_1(X_{1,1}) \le \frac{x_1}{t} \text{ or } 1 - F_2(X_{1,2}) \le \frac{x_2}{t}\right)\\
&= \lim_{t\to\infty} t\mathbb{P}\left(t\overline{F_1}(X_{1,1}) \le x_1 \text{ or } t\overline{F_2}(X_{1,2}) \le x_2\right)\\
&= \lim_{t\to\infty} t\mathbb{P}\left(t\left(\overline{F_1}(X_{1,1}), \overline{F_2}(X_{1,2})\right) \in \left([x_1, \infty] \times [x_2, \infty]\right)^c\right)\\
&= \int_0^{\pi/2}\left(\frac{x_1}{1 \vee \cot\varphi} \wedge \frac{x_2}{1 \vee \tan\varphi}\right)\Phi(d\varphi),\ x_1, x_2 \ge 0.
\end{aligned}$$

Formally, we have for any Borel set $\mathbf{A}$ in $[0,\infty]^2 \setminus \{(\infty,\infty)\}$, a measure Λ on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ defined by

$$\lim_{t\to\infty} t\mathbb{P}\left(\left(1-F_1\left(X_{1,1}\right), 1-F_2\left(X_{1,2}\right) \in At^{-1}\right)\right) = \Lambda\left(\mathbf{A}\right) \tag{3.5.3}$$

with $\Lambda\left(\partial\mathbf{A}\right) = 0$ (where $\partial\mathbf{A} = \mathbf{A} \setminus \mathbf{A}^c$), which is equivalent to

$$\Lambda\left(\left([x_1,\infty]\times[x_2,\infty]\right)^c\right) = \int_0^{\pi/2}\left(\frac{x_1}{1\vee\cot\varphi}\wedge\frac{x_2}{1\vee\tan\varphi}\right)\Phi\left(d\varphi\right),\ x_1,x_2\geq 0.$$

The relation (3.5.3) shows us how we can estimate the measure Λ from F, hence an intuitive estimator for F can be calculated by their empirical counterparts $F_{nj}(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(-\infty,x]}\left(X_{i,j}\right)$, which yields to the following estimator of Λ.

$$\tilde{\Lambda}_{t/n,n}\left(\mathbf{A}\right) = t\mathbb{P}_n\left(\frac{t}{n}\left(F_{n1},F_{n2}\right)\in\mathbf{A}\right) = \frac{t}{n}\sum_{i=1}^n \mathbf{1}_{\mathbf{A}}\left\{\frac{t}{n}\left(F_{n1},F_{n2}\right)\right\},$$

where $\mathbf{1}_{\mathbf{A}}$ is the indicator function for the events belong to $\mathbf{A}$. Writing $\tau = t/n$ we get

$$\tilde{\Lambda}_{\tau,n}\left(\mathbf{A}\right) = \tau\sum_{i=1}^n \mathbf{1}_{\mathbf{A}}\left(\tau\left(F_{n1},F_{n2}\right)\right).$$

Moreover, the extension of this measure to higher dimensions is straightforward. Consider that the limiting distribution of normalized componentwise maxima of $\mathbf{X}$ converges to the maximum domain of attraction of a multivariate extreme distribution. Furthermore, define the Borel set $[0,\infty]^d \setminus \{\infty\}$, then the measure Λ can be estimated directly by

$$\Lambda\left(\mathbf{A}\right) = \lim_{t\to\infty} t\mathbb{P}\left(t\left(\overline{F_1}\left(X_{1,1}\right),\ldots,\overline{F_d}\left(X_{1,d}\right)\right)\in\mathbf{A}\right) \tag{3.5.4}$$

and

$$\Lambda\left(\left([x_1,\infty]\times\cdots\times[x_d,\infty]\right)^c\right) = \int_{\mathbb{S}^{d-1}} w_j x_j \mathcal{H}(d\mathbf{w}),$$

which has the property

$$\int_{\mathbb{S}^{d-1}} w_j\mathcal{H}(d\mathbf{w}) = 1, \qquad j = 1,\ldots,d \tag{3.5.5}$$

for any Borel set $\mathbf{A}\subset[0,\infty]^d\setminus\{\infty\}$ with $\Lambda\left(\partial\mathbf{A}\right) = 0$ and

$$\mathbb{S}^{d-1} = \left\{(w_1,\ldots,w_{d-1}) : \sum_{j=1}^{d-1} w_j \leq 1,\ w_j\geq 0,\ j = 1,\ldots,d-1\right\}$$

is the $(d-1)$ dimensional unit simplex in $\mathbb{R}^d$. Further, $\mathcal{H}$ induces a positive finite measure on $\mathbb{S}^{d-1}$.

In practical terms, the empirical estimator is written as follows

$$\tilde{\Lambda}_{\tau,n} = \tau\sum_{i=1}^n \mathbf{1}_{\mathbf{A}}\left(\tau\left(F_{n1},\ldots F_{nd}\right)\right). \tag{3.5.6}$$

This simple estimator was introduced by Hsing et al. (2004) to estimate tail dependence in higher dimensions. We will use this estimator to derive other measures of dependence.

3.5.1. Characterization of the new measure of extreme dependence. We begin with the equations (3.5.4) and (3.5.6). In order to make inference in the dependence of extremal events we need to define the set $\mathbf{A}$ in a determining class of Λ. A set with such characteristics is

$$\mathbf{D}_{\varphi_1,\ldots,\varphi_{d-1}} := \left\{(x_1,\ldots,x_d) \in [0,\infty]^d : \ x_1 \wedge x_2 \tan\varphi_1 \wedge \cdots \wedge x_d \tan\varphi_{d-1} \leq 1\right\},$$

which has a clear geometric interpretation. This set was defined by Hsing et al. (2004) to measure tail dependence in a multivariate framework. For ease of the notation set

$$\Lambda(x_1,\ldots,x_d) := \Lambda\left(([x_1,\infty] \times \cdots \times [x_d,\infty])^c\right),$$

replacing $\mathbf{D}_{\varphi_1,\ldots,\varphi_{d-1}}$ in Λ we obtain

$$\psi(\varphi_1,\ldots,\varphi_{d-1}) := \Lambda\left(\mathbf{D}_{\varphi_1,\ldots,\varphi_{d-1}}\right) = \Lambda(1, \cot\varphi_1, \ldots \cot_{d-1}). \tag{3.5.7}$$

We will call $\psi(\varphi_1,\ldots,\varphi_{d-1})$ the *tail measure.*

The empirical counterpart of this measure of dependence is obtained by

$$\tilde{\psi}_{\tau,n}(\varphi_1,\ldots,\varphi_{d-1}) = \tau \sum_{i=1}^{n} \mathbf{1_A}\left(F_{n1} \leq \tau^{-1} \text{ or } \bigcup_{j=2}^{d} F_{n2} \leq \tau^{-1}\cot\varphi_{j-1}\right), \tag{3.5.8}$$

where $F_{n1} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,x]}(X_{i,j})$.

Following (3.5.8), one can deduce that the good results of the estimator are conditioned by the election of τ. Some properties of this function, which can be of great help on the context of estimation, are resumed in the next proposition.

PROPOSITION 3.5.1. *For all $\varphi_j \in [0,\pi/2]$ we have*

(1) The tail function $\psi(\varphi_1,\ldots,\varphi_{d-1})$ is convex and bounded by $\psi_{dep}(\varphi_1,\ldots,\varphi_{d-1}) \leq \psi(\varphi_1,\ldots,\varphi_{d-1}) \leq \psi_{ind}(\varphi_1,\ldots,\varphi_{d-1})$ where

$$\psi_{dep}(\varphi_1,\ldots,\varphi_{d-1}) = 1 \vee \cot\varphi_1 \vee \cdots \vee \cot\varphi_{d-1}$$

and

$$\psi_{ind}(\varphi_1,\ldots,\varphi_{d-1}) = 1 + \cot\varphi_1 + \cdots + \cot\varphi_{d-1}$$

(2) $\alpha\psi(\varphi_1,\ldots,\varphi_{d-1}) = \psi(\alpha\varphi_1,\ldots,\alpha\varphi_{d-1})$ for all $\alpha > 0$.

(3) for all $a_j > 0$ with $j = 1,\ldots,d-1$ we have

$$\left(1 \wedge \bigwedge_{j=1}^{d-1} a_j\right)\psi(\varphi_1,\ldots,\varphi_{d-1}) \leq \psi(a_1\varphi_1,\ldots,a_{d-1}\varphi_{d-1}) \leq \left(1 \vee \bigvee_{j=1}^{d-1} a_j\right)\psi(\varphi_1,\ldots,\varphi_{d-1}).$$

From (3.5.8) we observe that the difficulty in choosing τ is equivalent to the choice of a high threshold in classical extreme value theory. Since the homogeneity property 2 in Proposition 3.5.1, in the empirical section we pick τ so that $\psi_{\tau,n}$ mimics this property for a parameter $u > 0$. So for a fixed τ, we graph

$$\left\{\frac{\tilde{\psi}_{\tau,n}(u\varphi_1,\ldots,u\varphi_{d-1})}{u\tilde{\psi}_{\tau,n}(\varphi_1,\ldots,\varphi_{d-1})},\ u > 0\right\}. \tag{3.5.9}$$

The idea is that the ratio should be roughly constant and close to 1 for any $u > 0$, when τ is chosen adequately. The plots will look different for various values of τ, we will choose therefore τ for which the homogeneity property is more evident.

Now we give some results for the case that the subexponential property is present. A distribution function F on the non negative real line is called subexponential if $\lim_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n$ for all $n \geq 2$, where F^{*n} denotes the n-fold Stieltjes convolution of F with itself.

DEFINITION 3.5.2. Let $\mathbf{X} := \{X_1, \dots X_d\}$ be a d-variate vector of iid non-negative random variables with distribution F in $\mathbb{R}^d$, with $F(x) < 1$. The multivariate distribution is subexponential $\left(F \in S\left(\mathbb{R}^d\right)\right)$ if and only if

$$\lim_{t\to\infty} \frac{\overline{F^{*2}}(t\mathbf{x})}{F(t\mathbf{x})} = 2 \text{ for all } \mathbf{x} > 0 \text{ with } \bigvee_{j=1}^{d} x_j < \infty$$

and the iid copies of X_j satisfy the next relations

$$\mathbb{P}(S_{nj} > x_j) \sim n\mathbb{P}(X_j > x) \sim \mathbb{P}(M_{nj} > x_j)$$

as $x_j \to \infty$for $n = 2, 3, \dots$, and $j = 1, \dots, d$, where $S_{nj} = \sum_{i=1}^{n} X_{ij}$ and $M_{nj} = \bigvee_{i=1}^{n} X_{ij}$.

Note that multivariate subexponentiality implies multivariate regular variation. A textbook treatment of subexponential distributions is given in Embrechts et al. (1997); Kluppelberg (1998). Results in multivariate subexponetiality can be found in Omey et al. (2006); Omey (1994).

Some interesting properties of the tail function ψ on relation to this class of distributions is resumed as follows.

PROPOSITION 3.5.3. *Suppose that* $\mathbf{X} := \{X_1, \dots X_d\}$ *is a* d*-variate vector of iid non-negative random variables with multivariate subexponential distribution* $F \in S\left(\mathbb{R}^d\right)$. *Further, denote the notation* $\psi_{A,B}$ *the tail function between the components* A *and* B. *Then, as* $x_1, x_2 \to \infty$,

(1) for an iid non-negative random variable $X_1 = X_{11}, \dots X_{n1}$ with $F_1 \in S$, such that $\frac{\overline{F}(x_2)}{\overline{F}(x_1)} \to \cot\varphi_1$, we have

$$\psi_{M_{jn}, X_{j,n+1}}(\varphi_1) \sim (n-1) + \psi_{X_{j,1}, X_{j,n+1}}(\varphi_1),$$

(2) for two iid non-negative random variables $X_{11}, \dots X_{n1}$ and $X_{12}, \dots X_{n2}$ with $F_1, F_2 \in S$

$$\psi_{S_{n1}, S_{n2}}(\varphi_1) \sim n\psi_{X_{n1}, X_{n2}}(\varphi_1)$$

(3) Let Z and $X_1 \in \mathbf{X}$ be two independent random variables, with subexponential distributions $G, F_1 \in S$, such that $\mathbb{P}(X_j > x) = o(\mathbb{P}(X_1 > x))$ and $\frac{\mathbb{P}(Z > x_2)}{\mathbb{P}(X_{11} > x_1)} \to \cot\varphi_1$ for all $j \neq k$. Then

$$\psi_{S_{n1}, Z}(\varphi_1) \sim \psi_{X_1 Z}(\varphi_1).$$

(4) Define an iid non-negative random variable $X_1 = X_{11}, \dots X_{n1}$ with distribution $F_1 \in S$ and regular varying tail function $\overline{F}$ with index $-\alpha$, and $\mu_1, \dots \mu_n$ and $\upsilon_1, \dots, \upsilon_n$ non-negative constants. Further, Let $\chi = \sum_{i=1}^{n} \mu_i X_{i1}$ and $\Upsilon = \upsilon_i X_{i1}$ be two combinations of random variables with distributions $G_\chi, G_\Upsilon \in S$. then the

tail function is approximated by

$$\psi_{\chi\Upsilon}(\varphi_1) \sim \left(\sum_{i=1}^{n} \left(\mu_i^{-1} \wedge \upsilon_i^{-1} \cot\varphi_1\right)^{-\alpha} \right).$$

(5) Let $X_{11}, \ldots X_{n1}$ and $X_{12}, \ldots X_{n2}$ be iid non-negative regular varying random variables with index α, $\mu_1, \ldots \mu_n$ and $\upsilon_1, \ldots, \upsilon_n$ non-negative constants. Furthermore, define $\chi_1 = \sum_{i=1}^{n} \mu_i X_{i1}$ and $\chi_2 = \sum_{i=1}^{n} \upsilon_i X_{i2}$ with distributions $G_{\chi_1}, G_{\chi_2} \in S$. Then

$$\psi_{\chi_1\chi_2}(\varphi_1) = \frac{\mathbb{P}(\chi_1 > x_1 \text{ or } \chi_2 > x_2)}{\mathbb{P}(X_1 > x_1)} \sim \sum_{i=1}^{n} \mu_i^{\alpha} + \sum_{i=1}^{n} \upsilon_i^{\alpha} \cot\varphi_1.$$

Hsing et al. (2004) introduce a new dependence function which allows us to capture the complete extreme dependence structure and to present a nonparametric estimation procedure. The main advantage of this new dependence function is the possibility of visualization of extreme tail dependence in different directions.

Definition 3.5.4. (Hsing Definition 4.2) Let $\mathbf{X} := \{X_{n1}, \ldots X_{nd}\}$ be a d-variate vector of iid non-negative random variables with distribution $F \in \mathbb{R}^d$, and thresholds $x_1, \ldots, x_d \to \infty$ such that $\frac{\overline{F}(x_j)}{\overline{F}(x_1)} \to \cot\varphi_{j-1}$for all $j = 2, \ldots, d$, The $d-$variate classical tail dependence function λ is defined as follows

$$\lambda_d = \lim_{x\to\infty} \frac{\mathbb{P}\left(\bigcap_{j=1}^{d} X_j > x\right)}{\mathbb{P}(X_1 > x)}. \tag{3.5.10}$$

We define a new measure of tail dependence in terms of the dependence measure ψ. We call this measure the functional tail dependence and it is defined as follows

$$\rho(\varphi_1, \ldots, \varphi_{d-1}) = \frac{1 + \sum_{j=1}^{d-1} \cot\varphi_j - \psi(\varphi_1, \ldots, \varphi_{d-1})}{1 + \sum_{j=1}^{d-1} \cot\varphi_j - \bigvee_{j=1}^{d-1} \cot\varphi_j}, \text{ for } \varphi_j \in [0, \pi/2] \tag{3.5.11}$$

where $1 + \sum_{j=1}^{d-1} \cot\varphi_j$ and $\bigvee_{j=1}^{d-1} \cot\varphi_j$ are the tail functions ψ in the complete independent and dependent case. The limits of the new tail dependence function ρ are 0 or 1 corresponding to weak and complete dependence, respectively. Note, that when the functional tail dependence ρ is evaluated in $\pi/4$ it is equivalent to the classical measure λ.

The proof of this definition can be found in the Appendix of Hsing et al. (2004). Note that $\rho(\varphi)$ extends the definition of extreme tail dependence from a single direction, to all the directions. Moreover, $\rho(\varphi)$ is a copula property, hence it is invariant under monotone transformation of the marginal distributions. In the practical application we replace $\psi(\varphi_1, \ldots, \varphi_{d-1})$ by its empirical counterpart $\tilde{\psi}_n(\varphi_1, \ldots, \varphi_{d-1})$.

EXAMPLE 3.5.5. (The tail dependence function of M_4 processes). Let $\mathbf{X} = X_{n,j} = \bigvee_{l\in\mathbb{Z}^+}\bigvee_{\infty<k<-\infty}\alpha_{lkj}Y_{l,n-k}$ be a stochastic process, for $j=1,\ldots,d$, α_{lkj} non negative constants satisfying $\sum_{l\in\mathbb{Z}^+}\sum_{\infty<k<-\infty}\alpha_{lkj}=1$ and Y_{ln} is a vector of iid unit Frèchet random variables. Notice that

$$\begin{aligned}
&\mathbb{P}\left(X_{i,1}>x \text{ or } \bigcap_{j=2}^{d} X_{i,j}>x\tan\varphi_{j-1}\right)\\
&=1-\prod_{l\in\mathbb{Z}^+}\prod_{\infty<k<-\infty}\mathbb{P}\left(Y_{l,n-k}\le x\left(\alpha_{l1k}^{-1}\wedge\alpha_{l2k}^{-1}\tan\varphi_1\wedge\cdots\wedge\alpha_{ldk}^{-1}\tan\varphi_{d-1}\right)\right)\\
&=1-\exp\left\{-\frac{1}{x}\sum_{l\in\mathbb{Z}^+}\sum_{\infty<k<-\infty}\alpha_{l1k}\vee\alpha_{l2k}\cot\varphi_1\vee\cdots\vee\alpha_{ldk}\cot\varphi_{d-1}\right\}.
\end{aligned}$$

Furthermore, note that the limit of $\mathbb{P}\left(\bigcap_{j=1}^{d}X_{i,j}>x_j\right)/\mathbb{P}\left(X_{i,1}>x_1\right)\to\psi\left(\varphi_1,\ldots,\varphi_{d-1}\right)$ for $x_j\to\infty$. Hence,

$$\psi\left(\varphi_1,\ldots,\varphi_{d-1}\right)=\lim_{x\to\infty}\frac{1-\exp\left\{-\frac{1}{x}\sum_{l\in\mathbb{Z}^+}\sum_{\infty<k<-\infty}\alpha_{l1k}\vee\alpha_{l2k}\cot\varphi_1\vee\cdots\vee\alpha_{ldk}\cot\varphi_{d-1}\right\}}{\exp\left\{-\frac{1}{x}\right\}}$$

and by L'hopital we obtain

$$\psi\left(\varphi_1,\ldots,\varphi_{d-1}\right)=\sum_{l\in\mathbb{Z}^+}\sum_{\infty<k<-\infty}\alpha_{l1k}\vee\alpha_{l2k}\cot\varphi_1\vee\cdots\vee\alpha_{ldk}\cot\varphi_{d-1}.$$

Replacing the last result in 3.5.10 we obtain the d-variate functional tail dependence for a M_4 process.

$$\rho\left(\varphi_1,\ldots,\varphi_{d-1}\right)=\frac{1+\sum_{j=1}^{d-1}\cot\varphi_j-\sum_{l\in\mathbb{Z}^+}\sum_{\infty<k<-\infty}\alpha_{l1k}\vee\alpha_{l2k}\cot\varphi_1\vee\cdots\vee\alpha_{ldk}\cot\varphi_{d-1}}{1+\sum_{j=1}^{d-1}\cot\varphi_j-\bigvee_{j=1}^{d-1}\cot\varphi_j},\tag{3.5.12}$$

for $\varphi_j\in[0,\pi/2]$.

In the Example 3.3.1 we calculate the multivariate extremal index for a three dimensional case of a M_4 process. We generalize this example to show the three dimensional functional tail dependence of this process.

EXAMPLE 3.5.6. Consider the three dimensional M_4 process proposed in Example 3.3.1. We simulate $n=10.000$ realizations of $\mathbf{X}_n=\{X_{n,1},X_{n,2},X_{n,3}\}$, which are displayed in Figure 3.5.1. Applying (3.5.12) to this example we obtain the tail dependence function

$$\rho\left(\varphi_1,\varphi_2\right)=\frac{1+\cot\varphi_1+\cot\varphi_2-\left(\frac{1}{3}\vee\frac{\cot\varphi_1}{2}\vee\cot\varphi_2+\frac{1}{3}\vee\frac{\cot\varphi_1}{2}+\frac{1}{3}\right)}{1+\cot\varphi_1+\cot\varphi_2-\left(1\vee\cot\varphi_1\vee\cot\varphi_2\right)}.$$

Figure 3.5.2 shows the results of the simulation. The top panel displays the true tail dependence in three plots, a perspective plot, image plot and a contour plot. The bottom panel shows the same three plots for the estimated tail dependence function. The results are

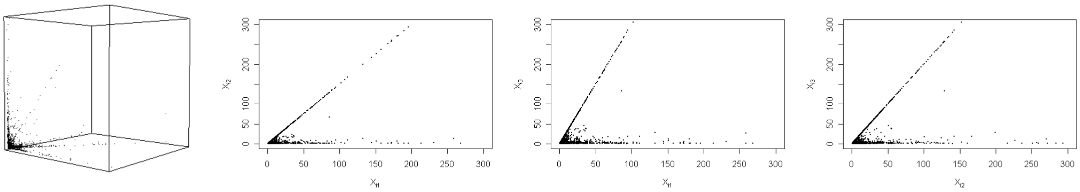

FIGURE 3.5.1. Simulated M_4 process for the Example 3.5.6 in 3-dimensional and with 2-dimensional projections

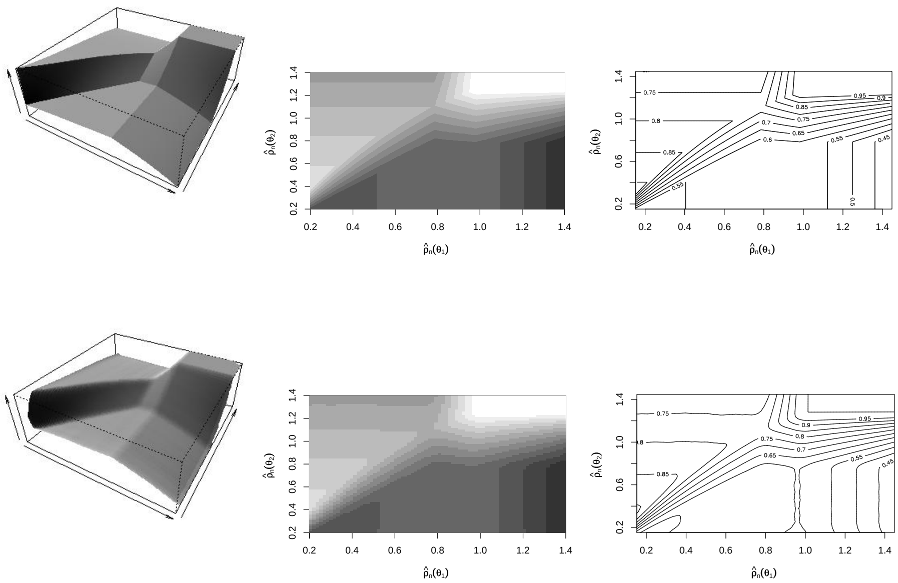

FIGURE 3.5.2. Simulation of the true trivariate functional tail dependence (top-panel) and estimated trivariate functional tail dependence function (low-panel) of the M_4 process in Example 3.5.6. The dark colors represent weak dependence ($\hat{\rho}_n(\varphi_1, \varphi_2) \to 0$) and the light color strong dependence ($\hat{\rho}_n(\varphi_1, \varphi_2) \to 1$).

quite indeed. For our sample we chose $\tau = 1/200$ and a smoothed version of $\hat{\rho}_n(\varphi_1, \varphi_2)$ by averaging this as follows

$$\hat{\rho}_n(\varphi_{i,1}, \varphi_{j,2}) = \frac{1}{(2k+1)^s} \sum_{u,v=-k}^{k} \hat{\rho}_n(\varphi_{i-u,1}, \varphi_{j-v,2}),$$

where $\varphi_{i,1}, \varphi_{j,2} \in \{\varphi_{w,\cdot} = w\pi/400\}$ for $w = 1, \ldots, 200$.

Note that if we had only estimated the Joe tail dependence function in this example we would obtain $\lambda = \rho(\pi/4, \pi/4) = 7/12$, which does not describe well enough the extreme dependence structure of this process.

The tail dependence as well as the functional tail dependence has the drawback that they are not a global measure of extreme dependence. Huang (1992) proposed a more general measure of tail dependence, *the conditional expected value of the probability of that k extreme events occur in d-marginals.* Let k be the number of extreme events that simultaneously occur in d-dimensions. In a finite sample framework, we have the following expression for the conditional expected value of the probability of that k extreme events occur, given that at least one $k \geq 1$ extreme event happened.

$$E\{k = d \mid k \geq 1\} \quad = \quad \frac{E\{k\}}{\mathbb{P}\{k \geq 1\}}.$$

This linkage measure has the advantages that it can be easily extended beyond the bivariate setting. Moreover, one does not need to condition on a specific marginal, whereby one would look only into one direction in the plane of extreme events.

For this reason, different authors have used this measure of extreme dependence. For instance, Straetmans (2000) used this measure to access extremal spillovers in equity markets. Hartmann et al. (2004) used it to investigate asset market linkages in crisis periods. De Vries (2005) to calculate the potential for systemic breakdowns in banks and finally, Geluk et al. (2007) proposed this measure as index of financial fragility.

However, we will demonstrate with an example that the maximum of this measure does not necessarily occur in the same threshold when $x_1 = \cdots = x_d \to \infty$. To demonstrate this we extend this measure to all the directions in terms of $\psi(\varphi_1, \ldots, \varphi_{d-1})$ as follows

$$\varepsilon_d(\varphi_1, \ldots, \varphi_{d-1}) = \frac{1 + \sum_{j=1}^{d-1} \cot \varphi_j}{\psi(\varphi_1, \ldots, \varphi_{d-1})}. \tag{3.5.13}$$

This measure has limits between 1 and d.

Example 3.5.7. Define the following three processes $X_1 = \xi_i$, $X_2 = \alpha\xi_i \vee (1-\alpha)\xi_{i+1}$ and $X_3 = \alpha\xi_i \vee (1-\alpha)\xi_{i+2}$, where ξ is iid random vector with Frèchet distribution and $0 < \alpha, \beta < 1$. Clearly the three processes are in some degree extreme dependent and it arises from the factors α and β. Moreover, note that X_1, X_2 and X_3 have Frechét marginals.

First, we calculate the tail function ψ for the trivariate case. We chose thresholds x_2, x_3 which satisfy the equation 3.5.7. This can be done by replacing $x_2 = x_1 \tan\varphi_1$ and $x_3 = x_1 \tan\varphi_2$. Then,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{3} X_i > x_i\right) &= 1 - \mathbb{P}(X_1 \leq x_1,\ X_2 \leq x_1 \tan\varphi_1,\ X_2 \leq x_1 \tan\varphi_2) \\
&= 1 - \mathbb{P}\big(\zeta_i \leq x_1 \wedge x_1\alpha^{-1}\tan\varphi_1 \wedge x_1\beta^{-1}\tan\varphi_2, \\
&\qquad \zeta_{i+1} \leq x_1(1-\alpha)^{-1}\tan\varphi_1,\ \zeta_{i+1} \leq x_1(1-\beta)^{-1}\tan\varphi_2\big) \\
&= 1 - \exp\left\{x_1^{-1}(1 \vee \alpha\cot\varphi_1 \vee \beta\cot\varphi_2\right. \qquad (3.5.14) \\
&\qquad \left. + (1-\alpha)\cot\varphi_1 + (1-\beta)\cot\varphi_2)\right\}.
\end{aligned}$$

Diving by $\mathbb{P}(X_1 > x_1)$ in (3.5.14) and evaluating in the limit when $x_1 \to \infty$ we obtain

$$\psi_{X_1X_2X_3}(\varphi_1, \varphi_2) = 1 \vee \alpha\cot\varphi_1 \vee \beta\cot\varphi_2 + (1-\alpha)\cot\varphi_1 + (1-\beta)\cot\varphi_2,$$

for all $\varphi_j \in [0, \pi/2]$.

Furthermore, we can derive the bivariate dependence function ψ for all the pairs between X_1, X_2 and X_3 through the trivariate case. Note, that $\psi_{X_1X_2}(\varphi_1) = \psi_{X_1X_2X_3}(\varphi_1, \pi/2)$ and

$\psi_{X_1X_3}(\varphi_2) = \psi_{X_1X_2X_3}(\pi/2, \varphi_2)$, which implies that

$$\begin{aligned}\psi_{X_1X_2}(\varphi_1) &= 1 \vee \alpha \cot \varphi_1 + (1-\alpha) \cot \varphi_1 \\ \psi_{X_1X_3}(\varphi_2) &= 1 \vee \beta \cot \varphi_2 + (1-\beta) \cot \varphi_2.\end{aligned}$$

For the case $\psi_{X_2X_3}$ we have to choose thresholds x_1, x_2, x_3 and divide by $\mathbb{P}(X_2 > x_2)$ in (3.5.14). From (3.5.7) is easy to show that $\psi_{X_1X_2X_3}(\varphi_1\varphi_2) = \psi_{X_1X_2X_3}(1, \varphi_1\varphi_2)$, but doing $\psi_{X_1X_2X_3}(0, 1, \varphi_2) = \psi_{X_2X_3}(\varphi_2)$, which means that we have to replace $x_1 = 0$ and $x_3 = x_2 \tan \varphi_2$ to obtain the wished result.

$$\psi_{X_2X_3}(\varphi_2) = \alpha \vee \beta \cot \varphi_2 + (1-\alpha) + (1-\beta) \cot \varphi_2.$$

Replacing these tail functions in (3.5.11) and (3.5.13) we can estimate the measure of the dependence ρ and ε respectively.

For example, we are interested in the functional tail dependence functions between the two pair of vectors (X_1, X_2) and (X_2, X_3) for $\alpha = \beta = 1/2$. Applying (3.5.11) and (3.5.13) we obtain that

$$\rho_{X_1,X_2}(\varphi) = \frac{1 \wedge 0.5 \cot \varphi}{1 \wedge \cot \varphi} \qquad \rho_{X_2,X_3}(\varphi_2) = \frac{1}{2}$$

$$\varepsilon_{X_1,X_2}(\varphi) = (1 + \tan \varphi) \wedge \left(1 + \frac{1}{2 \tan \varphi + 1}\right) \qquad \varepsilon_{X_2,X_3}(\varphi) = \frac{2(\cot \varphi + 1)}{1 \vee \cot \varphi + 1 + \cot \varphi}$$

for $\varphi\,[0, \pi/2]$.

Note that for these two bivariate pairs the classical tail dependence coefficient λ is equal for the two pairs $\lambda = \rho_{X_1,X_2}(\pi/4) = \rho_{X_2,X_3}(\pi/4)$. However, the functional tail dependences are quite different in other directions of the extremes! In fact, $\rho_{X_2,X_3}(\varphi) < \rho_{X_1,X_2}(\varphi)$ for $\varphi \in [0, \pi/4)$ and $\rho_{X_2,X_3}(\varphi) = \rho_{X_1,X_2}(\varphi)$ for $\varphi \in [\pi/4, \pi/2]$ (see Figures 3.5.3 and 3.5.4).

On the other hand, ε cannot distinguish the magnitude of the extreme events between the two pairs when $\varphi = \pi/4$. In particular $\varepsilon_{X_2,X_3}(\varphi) < \varepsilon_{X_1,X_2}(\varphi)$ for $\varphi \in [0, \pi/4)$, with the expected maximum value of $\varphi_{max} := \arctan(0.5)$ and $\varepsilon_{X_1,X_2}(\varphi_{max}) = 1.5$.

The tail dependence function ρ is the only one that takes its minimum value in $\varphi = \pi/4$, contrary to the measure ε, which can take its most informative or maximum value in other direction.

Thus, the classical tail dependence or the expected value measure, concentrate only on the same high thresholds, are not enough to characterize the extreme dependence between two or more random variables and therefore their implication in the works using these measures should be interpreted carefully.

We simulate $n = 10.000$ realizations of this example, the estimations of the two measures of extreme dependence for the three pairs (X_1, X_2, X_3) are shown in the Figures 3.5.5 and 3.5.6. The bottom panel shows the same three plots for the estimated measures of dependence. Observe that the results are nearly identical.

From a risk management point of view this should mean that the choice of investment portfolios is based only on the classical model of these measures, i.e., its evaluation in the diagonal of the first quadrant (in the bivariate case) might not be optimal from a diversification

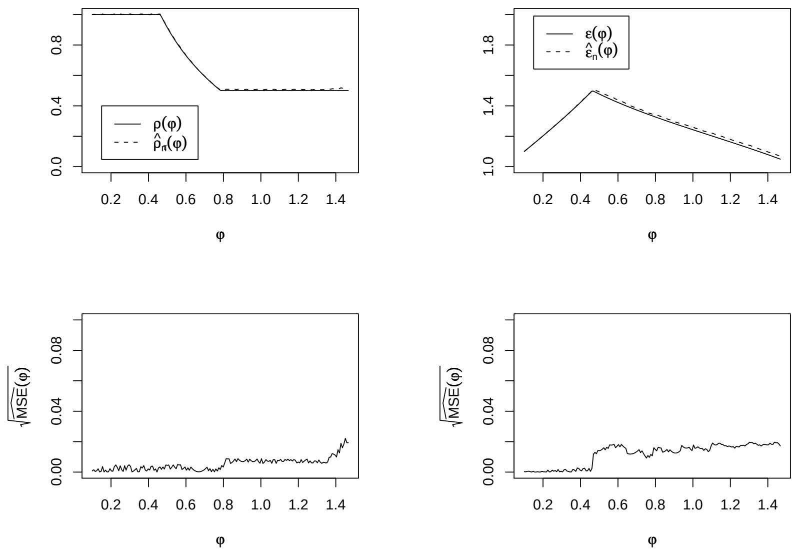

FIGURE 3.5.3. Emprical and theorical simulation of the measure of extreme dependences ρ(left column) and ε (right column) for the pair (X_1, X_2) in the Example 3.5.7 and its respective estimations of error in terms of $\sqrt{\widehat{MSE}}(\varphi)$ (bottom).

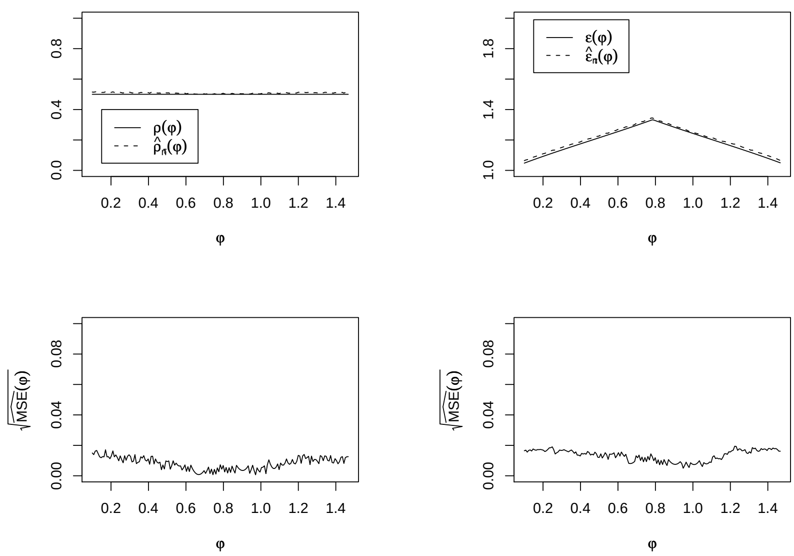

FIGURE 3.5.4. Emprical and theorical simulation of the measure of extreme dependence ρ (left column), δ (midle column) and ε (right column) for the pair (X_2, X_3) in the Example 3.5.7 and its respective estimations of error in terms of $\sqrt{\widehat{MSE}}(\varphi)$ (bottom).

perspective, while the measures proposed here in more general context should generate a better risk profile of diversification.

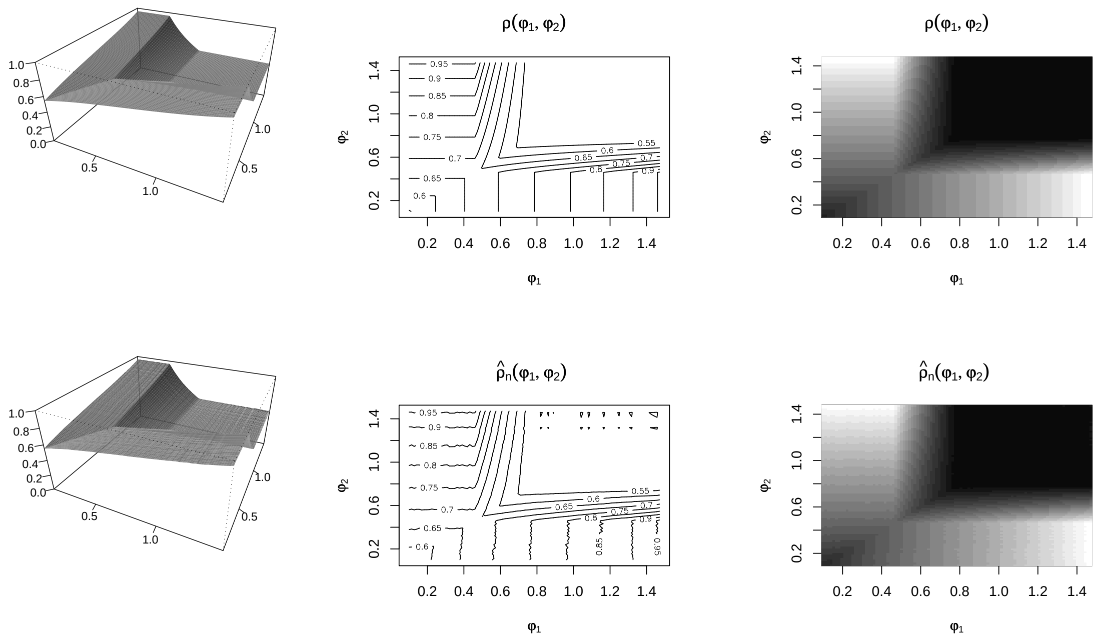

FIGURE 3.5.5. Theorical (top) and Empirical (bottom) simulation of thefunctional tail dependence measure ρ for the trivariate random vector (X_1, X_2, X_3) in the Example 3.5.7.

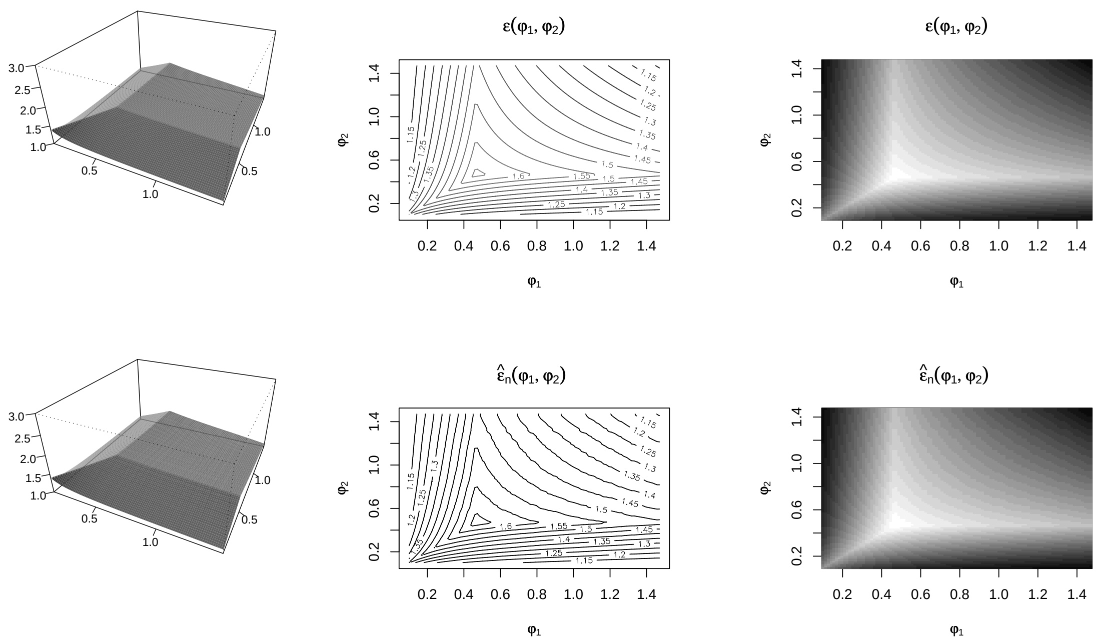

FIGURE 3.5.6. Theorical (top) and Empirical (bottom) simulation of the conditional expected crash probability measure ε for the trivariate random vector (X_1, X_2, X_3) of the Example 3.5.7.

3.6. Clusters and extreme dependence in the Asian crises

The set of results established in the last sections may provide reliable and practical inference for extremal events. To illustrate this possibility, this section is devoted to Asian financial markets.

Main issue in risk management is whether financial markets become more dependent during financial crisis or whether financial turbulence in emerging markets could endanger the stability of the global financial system. During the last decade the Asian equity markets have increasingly attracted non Asian investors, particularly from the U.S., with the aim to enjoy the benefit of diversification. However, extreme events as the Asian crisis suggest that the Asian capital markets have become increasingly integrated. The crisis first started in late 1997 in Thailand, then spread rapidly, causing turbulence in other East Asian financial markets, such as, Philippine, Malaysia, Indonesia and South Korea, while other authors claimed that turmoil was evident in equity markets earlier (see McKibbin and Martin (1998)).

Several studies have been concentrating on the possibility of contagion between these countries, where contagion is defined as a significant increase in cross-market linkages after a shock to one other country or group of countries. Other works of research have focused on establishing whether the Asian crisis had a contagious influence on the developed countries and vice versa. As such, the testable hypothesis of this section is to study the degree of cross-country stock market linkages before, during and after the crisis.

Whether or not the crisis period was characterized by contagion in Asian equity markets, it has attracted much attention. For example Arestis et al. (2005); Caporale et al. (2003); Bekaert and Harvey (2005) show some evidence concerning to contagion between Asian financial market. Much of the work on Asian financial market interrelationships has been constructed using correlation techniques, only few recent works have taken advantage of the sizeable advances in Copulas (see for example Caillault and Guegan (2005); Kole and Rotterdam (2006) and Rodriguez (2007)).

The main question in this section attempts to answer concerns about the likely effects of the bivariate extremal index, the cluster size probability and the tail dependence that may have had a log-run relationship with the Asian stock markets during the financial crisis. We focus on the equity markets of these countries since different authors suggest that the impact of contagion on return variation is more important for equity rather than currency markets (see Yang and Lim (2004) and Dungey et al. (2006)).

The sample consists of daily observations of stock index returns in local currency from Thailand, Philippine, Malaysia, Indonesia and South Korea. The hypothesis of different authors (Eichengreen et al. (1998); Dungey et al. (2006)) is that trade linkages, which are often between geographically proximate nations, are important for spillover effects. However, any of the five countries in discussion shares to other countries more than four percent of the total exports, making trade linkages an improbable source of extreme dependence. However, U.S.A and Japan are important clients of these exports, so that this argument could find some support. For this reason, we include a stock market from USA (S&P 500) and other big Asian market (Nikkei 225). All data are taken from Datastream.

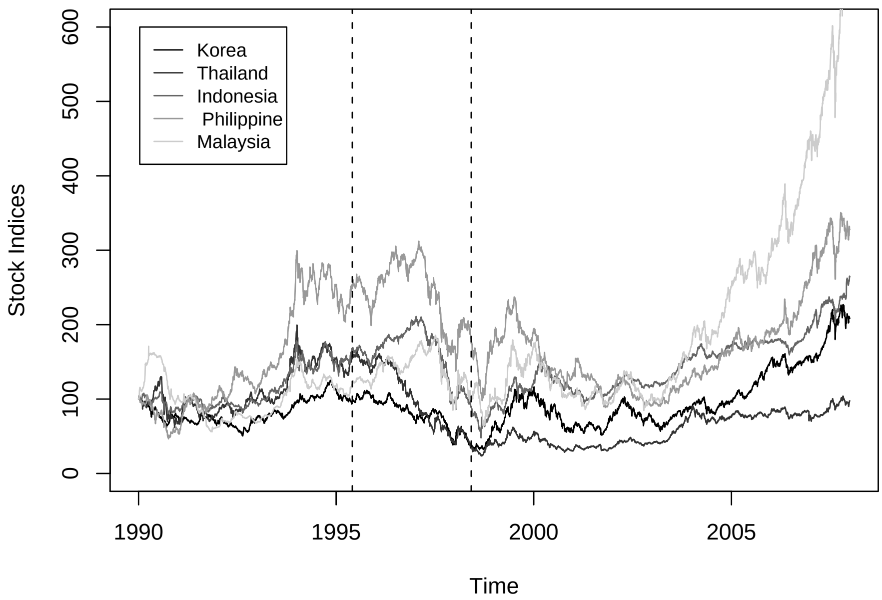

FIGURE 3.6.1. Stock market indices of Thailand, Philippine, Malaysia, Indonesia and South Korea. From January 1, 1990 to December 31, 2007.

In order to avoid the possible influence of other international crises as in Mexico in 1994 and in Russia in 1998, the sample was divided in three periods; from January 1, 1990 to May 31, 1995, from June 1, 1995 to May 31, 1998 and from June 1, 1996 to December 31, 2007. The full sample consists of 4695 daily return observations, which allow us to identify a sufficient number of extreme return observations to estimate statistical models for these rare events. The returns are calculated as the negative log-difference of each series, so that we can work with the maxima of the sample and not the minima. In each period of analysis we consider the 0.97- quantile of the return distribution as the threshold for the definition of extremes events.

Different angles allow us to capture a better measure of extreme events, reflecting that, for example, the probability of markets crashing together is higher in periods of financial turmoil reflexes by means of the tail dependence function, while the MEI and the functional cluster size of probabilities can describe the cluster behaviour of these crashes.

3.6.1. Extreme dependence between Asian countries. Table 3.B.1 presents a wide range of descriptive statistics for each series under the three periods of investigation. The mean return is close to zero for all series. However, they differ considerably in terms of standard deviation, skewness and kurtosis. The assumption of normally distributed returns is strongly rejected by all series through the Jarque-Bera test. Other assumptions such as the null hypothesis that the returns series are iid random variables as well that the returns have a unit root are strongly rejected.

The first step towards an analytical investigation into the performance of the different markets during the Asian crisis as for example; the univariate extremal index and the probability of cluster size are presented in Table 3.B.2. An interesting result during the extreme events in the stock markets in the second period is that, the cluster size probability is higher for small number of size cluster extremes in comparison to other cluster sizes in the period before and after the crisis. Hence, the extremal index is larger with the exception of Malaysia.

In order to gain an insight into the nature of the relationship between the extreme events in stock markets, bivariate analysis were carried out first between the Asian stock markets. In relation to the widely reported experiences of linkages among Asian markets only two countries, Thailand and Malaysia, may be considered as the likely origins.

For this reason, we concentrate firstly on Thailand versus the other countries. The results are displayed in the Figures 3.B.1, 3.B.2, 3.B.3 and 3.B.4.

Focusing on the bivariate extremal index we observe that it grows in the crisis period in contrast to the pre and post crisis period for all countries. With the exception of Philippine and Malaysia the bivariate extremal index is constant for different values of $w \in [0, 1]$, i .e., the direction of the extremes. Furthermore, a more detailed analysis displayed in the low panel in each of the figures shows that the distribution of the cluster size probabilities for different sizes k are quite different. For the most of the countries the cluster size probability $\pi_n(w, 1)$ is the highest for all the periods, but especially during the crisis period. The exception is the case Thailand vs Philippine in the pre period crisis, where the probability $\pi_n(w, k)$ is circa of 0.6 and 0.2 for $k = 1$ and $k = 2$ when $w \to 0$, which signifies that the extreme events for Thailand tend to happen three times more frequently that cluster of extreme events of size two. For $w \to 1$, the probability $\pi_n(w, k)$ for $k = 1, 2$ is about 0.4 for both stock markets.

On the other hand, during the period of crisis the bivariate extremal index behaves close to a completely dependent bivariate extremal index for Thailand versus Philippine and Thailand versus Malaysia. Even more interesting is to observe the changes in the cluster size probability $\pi_n(w, k)$ for these countries. For $\lim_{w\to 0} \pi_n(w, k)$ we have that the probability of cluster of size one augments about 20% and probability of cluster of size two decreases of 15% in comparison to the pre-period of crisis. However, the cluster size probability augments also for cluster size bigger than three.

In the post-crisis period one can observe that, the bivariate extremal index is more or less constant for all $w \in [0, 1]$. In the same way the probability $\pi_n(w, k)$ does not show many variations between the cluster sizes, with the exception of $w \approx 0.5$ where $\pi_n(w, k)$ grows for $k = 2$ and diminishes for $k = 1$. The tail dependence for all the countries increases during time of crisis. When $\theta \sim \pi/4$ the difference between the tail dependence pre, post and crisis period is maximal, about 0.5 for all pairs of countries. This means, that for each two extreme events in the Thailand stock market index, we can expect an extreme event in some other countries.

In the case Philippine versus the other countries the results are displayed in the Figures 3.B.5,3.B.6 and 3.B.7. Three important characteristics can be obtained from these pictures. First, like the case Thailand vs Philippine, the bivariate extremal index changes in periods of the crisis and it is relatively higher than the pre and post period crisis, with the exception of Philippine vs Malaysia.

Second, the probability of cluster size $\pi_n(w,1)$ is always bigger for all the periods. However, the probability of cluster size $\pi_n(w,2)$, $\pi_n(w,3)$ and $\pi_n(w,4)$ vary sharply when w is approaching to 0.5 in the crisis period. Particularly,$\pi_n(w,4)$ increases in crisis period for Philippine versus Malaysia by about 10%. For Philippine versus Indonesia and Korea $\pi_n(w,3)$ augments in the same proportionality. Furthermore, for this three countries the probability of cluster size $\pi_n(w,2)$ diminishes during time of crisis. Unlike the analysis Thailand versus the other countries, the probability of cluster size $\pi_n(w,2)$ keeps the higher probability after the probability $\pi_n(w,1)$ for all countries during the post crisis period.

Finally, the functional tail dependence between these stocks market returns is higher for the crisis, pre and post period in decreasing order. The functional tail dependence is exceptionally marked for the case Philippine versus Indonesia during crisis period by about 0.25 higher.

The cases Malaysia versus Indonesia and Malaysia versus Korea are displayed in the Figures 3.B.8 and 3.B.9 and one can observe that the bivariate extremal index varies through the periods but without a defined pattern. Moreover, the bivariate extremal index for these two countries is very similar for the crisis and post period. However, deeper analyses in the cluster size probabilities reveal the most notorious difference. During the crisis period, the probability $\pi_n(w,1)$ is the highest for the two pairs of countries and they follow the same path for different values of w. Despite the fact that, the probabilities $\pi_n(w,3)$, $\pi_n(w,4)$ are divergent for distinct values of w. For example, the case of Malaysia versus Indonesia the probabilities $\pi_n(w,3)$ and $\pi_n(w,4)$ are relatively higher for values of w near of 0.5 during the crisis, implying that the extreme events tend to be of size three and four with a higher probability. Additionally, the tail dependence is more marked in time of crisis for the two pair of countries. On the other hand, the differences between the tail dependence in the pre and post crisis period are not distinguishable.

The last pair of countries in this analysis is Indonesia versus Korea. Figure 3.B.10 shows the main results. Like in the previous analysis the bivariate extremal index increases in the crisis periods. Paradoxically, the bivariate extremal index is the lowest for the post-crisis period. A more detailed analysis in the cluster size probabilities in the pre and post crisis periods reveals that the most important variation is in relation to the probability $\pi_n(w,1)$ by about 0.3 in the post crisis period. On the other hand during the crisis period, the probabilities $\pi_n(w,1)$ and $\pi_n(w,2)$ increases and diminishes respectively, for values near $w=0.5$, i.e., for extreme values that are situated in the diagonal. The final result is related to the functional tail dependence of the data. During pre and post crisis period the dependence is similar, while in the crisis period the tail dependence is intensified by about 0.2.

3.6.2. Extreme dependence between the Asian-5 countries, U.S.A and Japan stock markets. We investigated extreme stock market dependence between U.S.A, Japan and Asian markets. Japan did not suffer from a crisis, though stock prices in Japan are likely to influence other Asian stock markets and hence are also considered. Using the multivariate extremal index and the functional tail dependence, we find the following empirical results. First, there exists significant increase in extreme events during the period of crisis. This contagion effects are found in the Japan and U.S.A market. Second, contagion effects defined as an increment in the functional tail dependence are stronger from the U.S.A market to Asian markets than from Japan, indicating that U.S. market plays a major role in the transmissions of information to foreign markets. Third, the intensity of contagion is significantly greater during the Asian financial crisis than after the crisis.

All of these results appear sensible in terms of the relative importance of these markets in the Asian region. One of the most interesting findings concerns the change in the most influential markets, as measured by the functional tail dependence, in the post-crisis period compared to the pre-crisis period. Once again, this is largely consistent with the notion of "contagion effects" following the onset of the Asian crises and the greater degree of market interdependence in the post-crisis regional economy. Overall, these findings are comparable to most other works in this area. The results obtained in this chapter complement this work in quantifying the interdependencies among Asian equity markets.

3.7. Conclusions

In brief, we have contributed to the literature in two ways. First, we have introduced a suitable framework to estimate the MEI and other cluster functionals. Specifically, we have adopted the Pickands representation.

Second, we have provided some empirical evidence of the transmission of turbulent crisis period across Asia and other two development countries, by applying the suggested methodology to daily data on stock returns, estimating bivariate models for the Asian countries and trivariate models for each Asian country and the other development countries.

In particular, we have examined whether during the 1997 East Asian crisis there was any contagion between Thailand, Indonesia, Philippine, Korea, Malaysia and the other two developed countries (Japan, U.S.A). We have tested contagion as a positive shift in the degree of comovement between asset returns, taking into account the multivariate extremal index, the cluster size probabilities and the tail dependence function. The estimation results show that the impact of the extreme dependence between the Asian countries was considerable. There is evidence of changing dependence structures during periods of financial turmoil.

Possible reasons include long-standing trends in trade and investment interaction, the more recent convergence in monetary policies and the almost universal process of microeconomic reform flowing from the crises themselves. A general conclusion is that the functional tail dependence is nearly symmetric for all countries investigated in this chapter. By contrast, in the normal periods, i.e., pre and post crisis, we do not find evidence of a significant tail dependence across stock markets for any pairs of Asian countries. This suggests that at least some markets have become more isolated following these macroeconomic shocks. The findings obtained in this chapter have obvious implications, amongst other things, for the purported benefits of international portfolio diversification among the several Asian equity markets. As a result, the strong linkages among the national markets would indicate that the returns from such a strategy have diminished markedly. However, the results also suggest that opportunities for diversification may still exist, especially in some of the smaller markets. In the pre-crisis period most Asian equity markets were relatively isolated from each other or were subject to only a few direct extreme events. One of the most interesting findings concerns the change in the most influential markets, as measured by the functional tail dependence, in the post-crisis period as compared to the pre-crisis period. This is at least one indication of an increasingly interdependent Asian regional market. These results contribute to the ongoing debate on the existence of contagion from the point of view of the experimented extreme events.

3.A. Demonstrations

PROOF. *Theorem 3.2.2.* Let l_n and m_n be positive integers such that $l_n = o(m_n)$, $m_n p_n^\tau \to 0$, $\max_{s=l_n,\ldots,r_n} \left\{\alpha^\tau_{s,l_n,n}\right\} = o(m_n p_n^\tau)$ for every $0 < x < \liminf_{n\to\infty} r_n p_n^\tau$ and $s_n \leq r_n - m_n$.

Theorem 5.7 and 5.9 in Seger (2002) implicate, that

$$\max\left\{s = m_n + l_n, \ldots, r_n - m_n : \left|\theta_{s,n}(\tau) - \theta_{m_n n, q_s^\tau, n}(\tau)\right|\right\} \to 0$$

and $q^\tau_{s_n,n} \leq (1 - \theta_{m_n,n}(\tau)\, p_n^\tau)^{s_n} + o(1) = \exp(-s_n \theta_{m_n,n}(\tau)\, p_n^\tau) + o(1)$. Then, if $s_n = \lceil x/p_n^\tau \rceil$ we have $\theta_{s_n,n}(\tau) = \theta_{m_n,n}(\tau)\, q^\tau_{s_n,n} + o(1)$ and $\theta_{s_n,n}(\tau) = \theta_{m_n,n}(\tau)\exp(-x\theta_{m_n,n}(\tau)) + o(1)$.

Furthermore, the characterization theorem in Seger (2002) shows that $\theta_{m_n,n}(\tau) = \theta_n(\tau) + o(1)$. Now, following Proposition 2.1 in Smith and Weissman (1996), we need to find thresholds $\mathbf{u}_n \in \mathbb{R}^d_+$, which satisfies

$$n\mathbb{P}(\mathbf{X_i} > \mathbf{u}_n) \to \tau. \tag{3.A.1}$$

Then, replacing $\mathbf{u}_n = n/\tau$ in (3.2.4) and $\mathbf{u_n} = n$ in (3.2.5) we obtain

$$\mathbb{P}(\chi_1(\tau) > n) = p_n^\tau = p_{S,n} \to v,$$

for $v > 0$ and

$\mathbb{P}\left(\bigvee_{i=1}^n \chi_i(\tau) \leq n\right) = \mathbb{P}\left(\bigvee_{i=1}^n X_{i,1} \leq u_{n1}, \ldots, \bigvee_{i=1}^n X_{i,d} \leq u_{nd}\right) = \mathbb{P}(\mathbf{M}_n \in S_n) \to \exp(-v\theta_n(\tau))\,.$

Hence, these limits converge when $n \to \infty$, which yields to

$$\mathbb{P}(p_{S_n} T_{S,n} > n) \to \mathbb{P}(p_n^\tau T_{B,n}(\tau) \geq n) = \theta_n(\tau)\exp(-n\theta_n(\tau)) + o(1)\,,$$

completing the proof. □

PROOF. *Proposition* 3.5.1 □

(1) For the demonstration of the Properties 1 see appendix in Hsing et al. (2004).

(2) Note that by definition

$$\mathbb{P}\left(\bigcup_{j=1}^d X_j > x_j\right) \sim \frac{1}{n}\Lambda\left(n\overline{F}(x_1), \ldots, n\overline{F}(x_d)\right) \tag{3.A.2}$$

for $n \to \infty$. Then, remplacing $t = n/\alpha$ in (3.A.2) such that $t \to \infty$. We have

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j=1}^d X_j > x_j\right) \quad &\sim \quad \frac{1}{t\alpha}\Lambda\left(t\alpha\overline{F}(x_1), \ldots, t\alpha\overline{F}(x_d)\right)\\ &= \alpha^{-1}\overline{F}(x_1) \quad \Lambda\left(\alpha, \alpha\frac{\overline{F}(x_2)}{\overline{F}(x_1)}, \ldots, \alpha\frac{\overline{F}(x_d)}{\overline{F}(x_1)}\right).\end{aligned} \tag{3.A.3}$$

On the other hand we have

$$\psi(\varphi_1, \ldots, \varphi_{d-1}) := \frac{\mathbb{P}\left(\bigcup_{j=1}^d X_j > x_j\right)}{\mathbb{P}(X_1 > x_1)}.$$

Now dividying by $\overline{F}(x_1)$ in (3.A.3) we obtain the wished result.

(3) Let be $a_0, \ldots, a_{d-1}$ a series of positive constants and for ease of the notation $\tan\varphi_0 = 1$ for all $\varphi_0 \in [0, \pi/2]$. Furthermore, let $a_{min} := 1 \wedge \bigwedge_{j=1}^{d-1} a_j$, then for (3.5.7) and

the Propertie 2

$$\begin{aligned}
\psi\left(a_1\varphi_1,\ldots,a_{d-1}\varphi_{d-1}\right) &= \lim_{t\to\infty} t\mathbb{P}\left(\bigcup_{k=0}^{k=d-1} t\overline{F_{k+1}}\left(X_{1,k+1}\right)\le a_k\tan\varphi_k\right)\\
&= \lim_{t\to\infty} a_{min}t\mathbb{P}\left(\bigcup_{k=0}^{k=d-1} t\overline{F_{k+1}}\left(X_{1,k+1}\right)\le \frac{a_k}{a_{min}}\tan\varphi_k\right)\\
&= a_{min}\psi\left(\frac{a_1}{a_{min}}\varphi_1,\ldots,\frac{a_{d-1}}{a_{min}}\varphi_{d-1}\right)\ge a_{min}\psi\left(\varphi_1,\ldots,\varphi_{d-1}\right).
\end{aligned}$$

The other bound can be demonstrated similarly.

PROOF. *Proposition* 3.5.3 □

(1) Because of the properties of independence and the subexponential distribution of X_j we have

$$\begin{aligned}
\mathbb{P}\left(M_{jn}>x_1 \text{ or } X_{j,n+1}>x_2\right) &\sim \overline{F^{*n}}(x_1)+\overline{F}(x_2)\\
&\sim n\overline{F}(x_1)+\overline{F}(x_2)
\end{aligned}$$

diving by $\overline{F}(x_1)$ yields to.

$$\psi_{M_{jn},X_{j,n+1}}\left(\varphi_1\right)\sim\frac{(n-1)\,\overline{F}(x_1)+\overline{F}(x_1)+\overline{F}(x_2)}{\overline{F}(x_1)}=(n-1)+\psi_{X_{j,1},X_{j,n+1}}\left(\varphi_1\right).$$

(2) By the multivariate subexponential definition we have

$$\begin{aligned}
\psi_{S_{n1},S_{n2}}\left(\varphi_1\right)\overline{F}(x_1) &= \mathbb{P}\left(S_{n1}>x_1 \text{ or } S_{n2}>x_2\right)\\
&\sim \overline{F^{*n}}(x_1)+\overline{F^{*n}}(x_2)-n\mathbb{P}\left(X_1>x_1\,,\ X_2>x_2\right)\\
&= n\left(\overline{F}(x_1)+\overline{F}(x_2)-\mathbb{P}\left(X_1>x_1\,,X_2>x_2\right)\right)
\end{aligned}$$

which conduces to $\psi_{S_{n1},S_{n2}}\left(\varphi_1\right)=n\psi\left(\varphi_1\right)$.

(3) By dominance of $\mathbb{P}\left(X_j>x\right)=o\left(\mathbb{P}\left(X_1>x\right)\right)$ for all $j=2,\ldots,d$ and independence $\mathbb{P}\left(S_{n1}>x_1 \text{ or } Z>x_2\right)\sim\overline{F^{*n}}(x_1)+\overline{F_Z}(x_2)$. We have that $\overline{F^{*n}}(x_1)\sim\overline{F}(x_1)$. Hence, $\psi_{S_{n1},Z}\left(\varphi_1\right)\sim\psi_{X_1Z}\left(\varphi_1\right)$.

(4) By regular variation $\mathbb{P}\left(\sum_{i=1}^n\mu_iX_{i1}>x_1\right)\sim\mathbb{P}\left(X_1>x_1\right)\sum_{i=1}^n\mu_i^\alpha$, i.e, $\overline{F_\chi}(x_1)\sim\sum_{i=1}^n\mu_i^\alpha\overline{F}(x_1)$ and $\overline{F_\Upsilon}(x_2)\sim\sum_{i=1}^n\upsilon_i^\alpha\overline{F}(x_2)$. Now by absolute dependence between χ and Υ we have that $\mathbb{P}\left(\chi>x_1\ \boldsymbol{\Upsilon}>x_2\right)=\sum_{i=1}^n\mu_i^\alpha\overline{F}(x_1)\wedge\sum_{i=1}^n\upsilon_i^\alpha\overline{F}(x_2)$, diving by $\overline{F}(x_1)$ the result follows.

(5) This demonstration follows the same arguments as the last proof, with the difference that χ_1 and χ_2 are now independent.

3.B. Tables and Figures

	Hong Kong	Korea	Thailand	Indonesia	Philippine	Malaysia
Obs.	4695	4695	4695	4695	4695	4695
Min	-14.735	-12.805	-16.063	-22.697	-9.744	-12.732
Max	17.247	10.024	11.350	17.248	16.178	13.128
Mean	0.049	0.016	-0.001	0.021	0.025	0.041
Std.Dev	1.538	1.835	1.705	1.349	1.527	1.475
Skewness	-0.048	-0.069	0.089	-0.136	0.453	0.171
Kurtosis	10.259	4.209	6.697	40.850	8.783	10.483
Phillips-Perron Unit Root Test	-4536*	-4266*	-4334*	-4358*	-3918*	-3809*
Jarque-Bera Test	20612*	3474*	8789*	326761*	15268*	21542*

Table 3.B.1. Summary statistics for the stock market returns. Asymptotic p-value are shown in the brackets. *,**,*** denote statistical significance at the 1, 5 and 10 % level respectively.

	extremal index	cluster sizes			
First period	θ	1	2	3	$\leq$4
Thailand	0.502 (0.492, 0.511)	0.63	0.18	0.00	0.05
Philippine	0.439 (0.412, 0.441)	0.33	0.39	0.05	0.23
Malaysia	0.516 (0.486, 0.532)	0.59	0.23	0.09	0.09
Indonesia	0.676 (0.662, 0.692)	0.75	0.11	0.07	0.07
Korea	0.651 (0.627, 0.651)	0.64	0.21	0.11	0.04
Second Period	θ	1	2	3	$\leq$4
Thailand	0.832 (0.821, 0.847)	0.85	0.10	0.05	0.00
Philippine	0.585 (0.567, 0.592)	0.50	0.36	0.07	0.07
Malaysia	0.381 (0.367, 0.413)	0.33	0.22	0.22	0.23
Indonesia	0.821 (0.820, 0.844)	0.65	0.28	0.06	0.01
Korea	0.845 (0.834, 0.855)	0.66	0.28	0.05	0.01
Third Period	θ	1	2	3	$\leq$4
Thailand	0.508 (0.496, 0.517)	0.59	0.21	0.00	0.20
Philippine	0.610 (0.597, 0.612)	0.57	0.25	0.02	0.16
Malaysia	0.650 (0.646, 0.659)	0.65	0.19	0.13	0.03
Indonesia	0.263 (0.252, 0.279)	0.45	0.20	0.05	0.10
Korea	0.390 (0.365, 0.408)	0.36	0.14	0.18	0.32

Table 3.B.2. Univariate extremal index estimation.

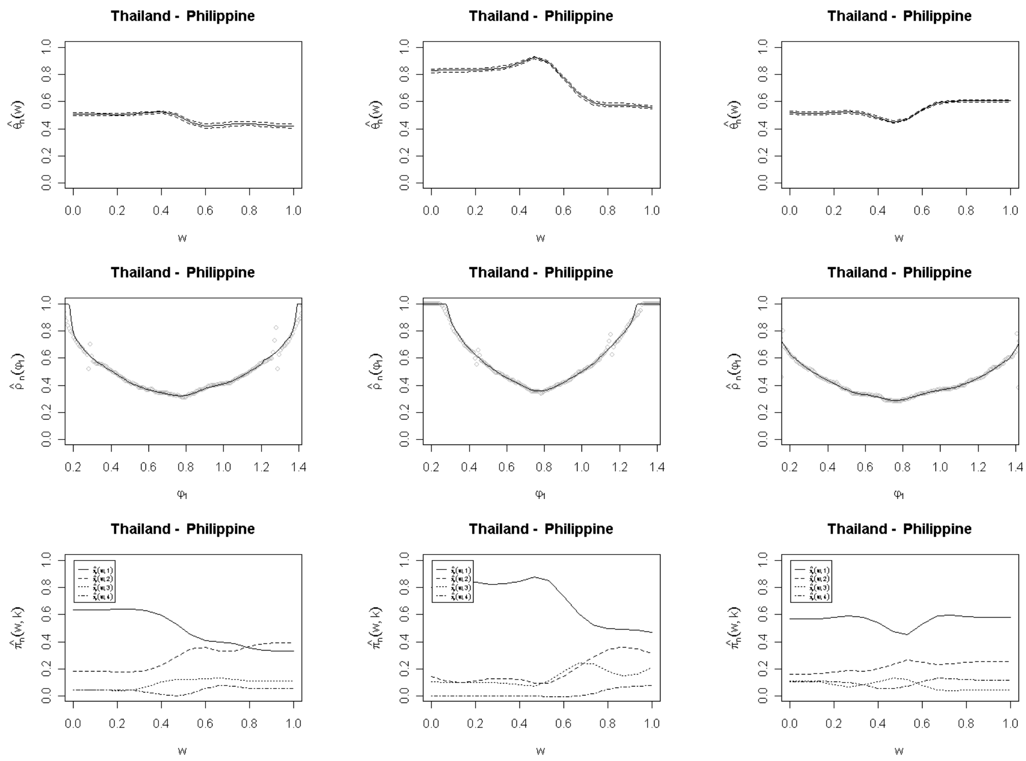

FIGURE 3.B.1. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Thailand vs Philippine

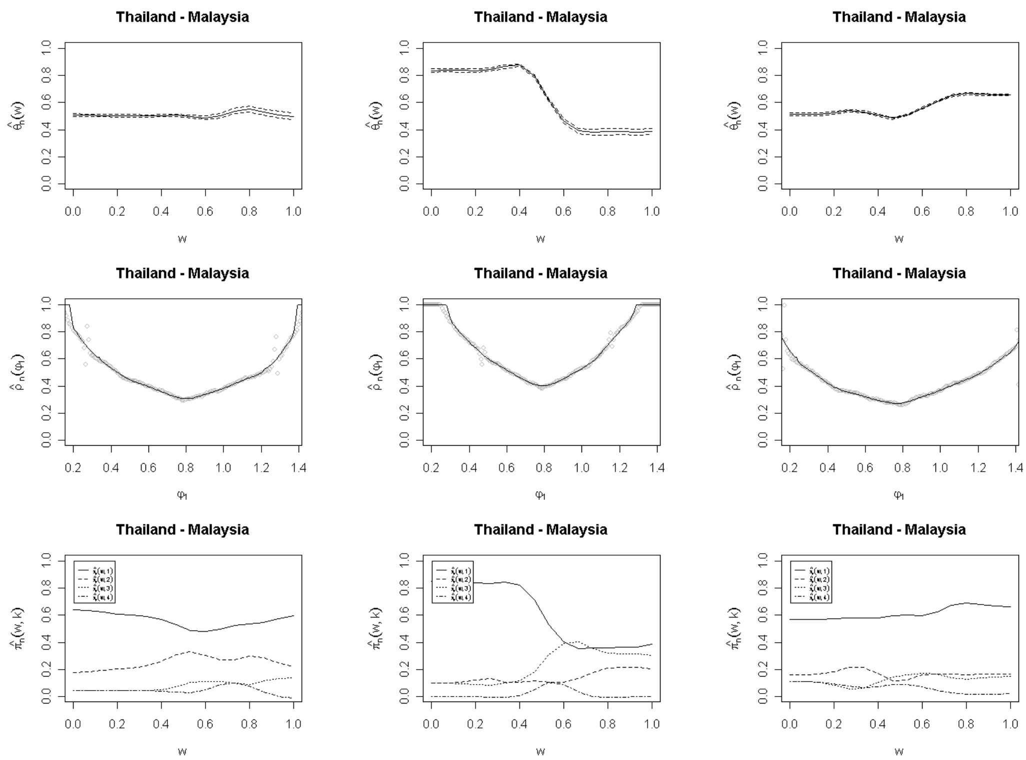

FIGURE 3.B.2. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Thailand vs Malaysia.

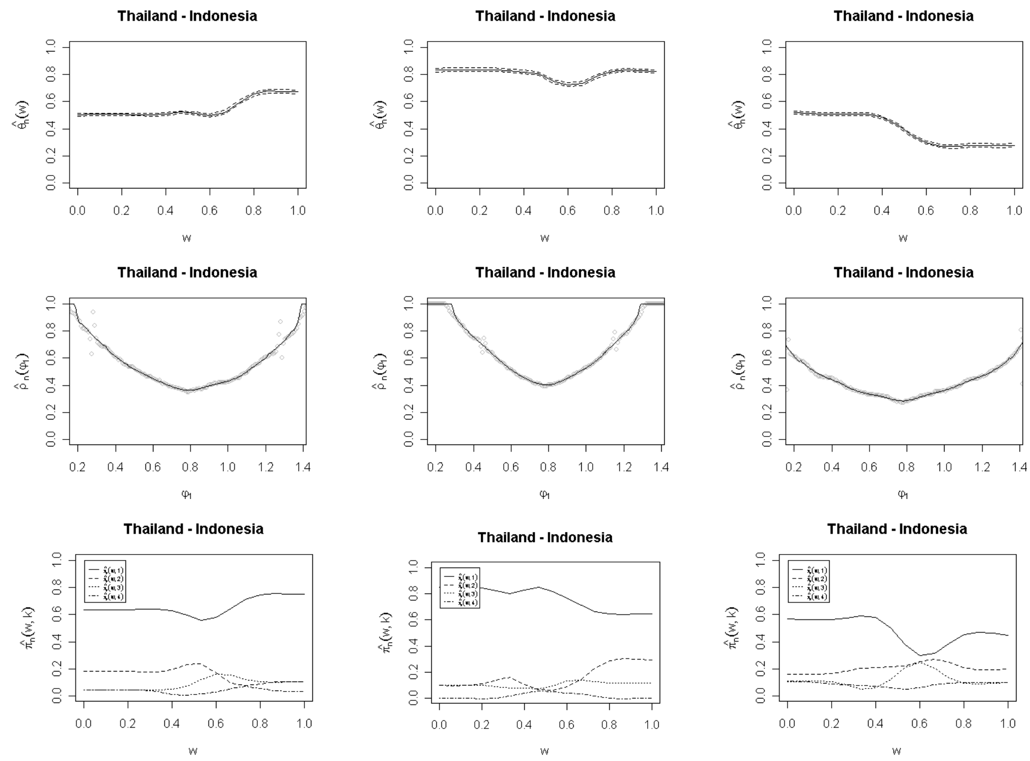

FIGURE 3.B.3. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Thailand vs Indonesia.

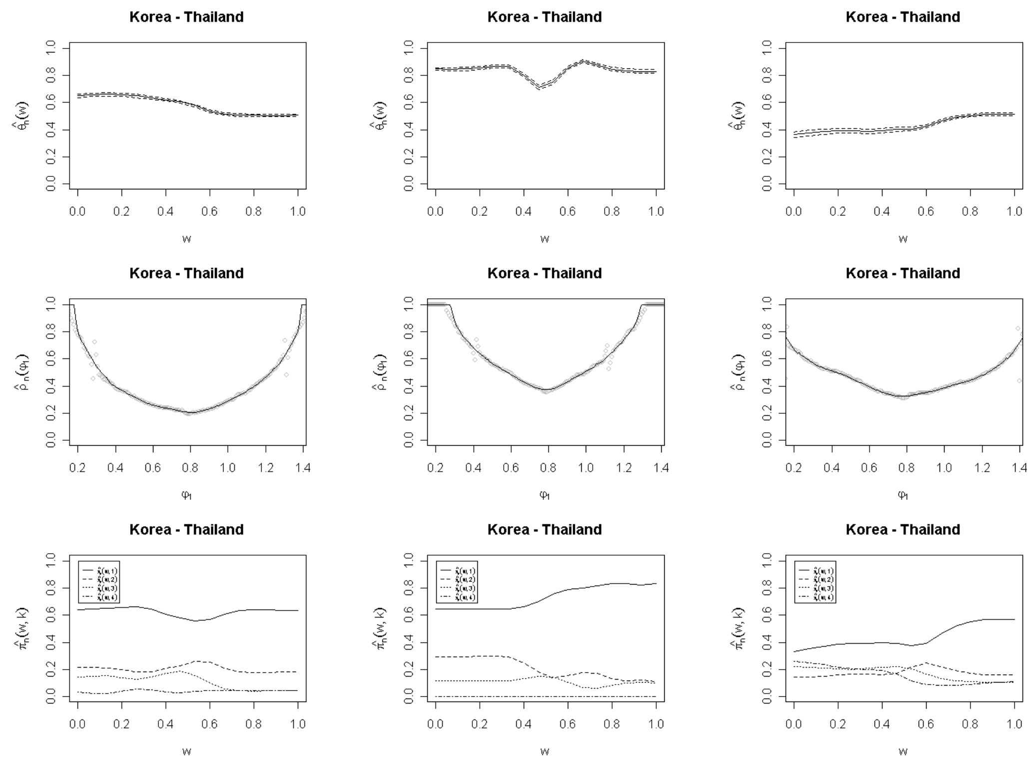

FIGURE 3.B.4. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Thailand vs Korea.

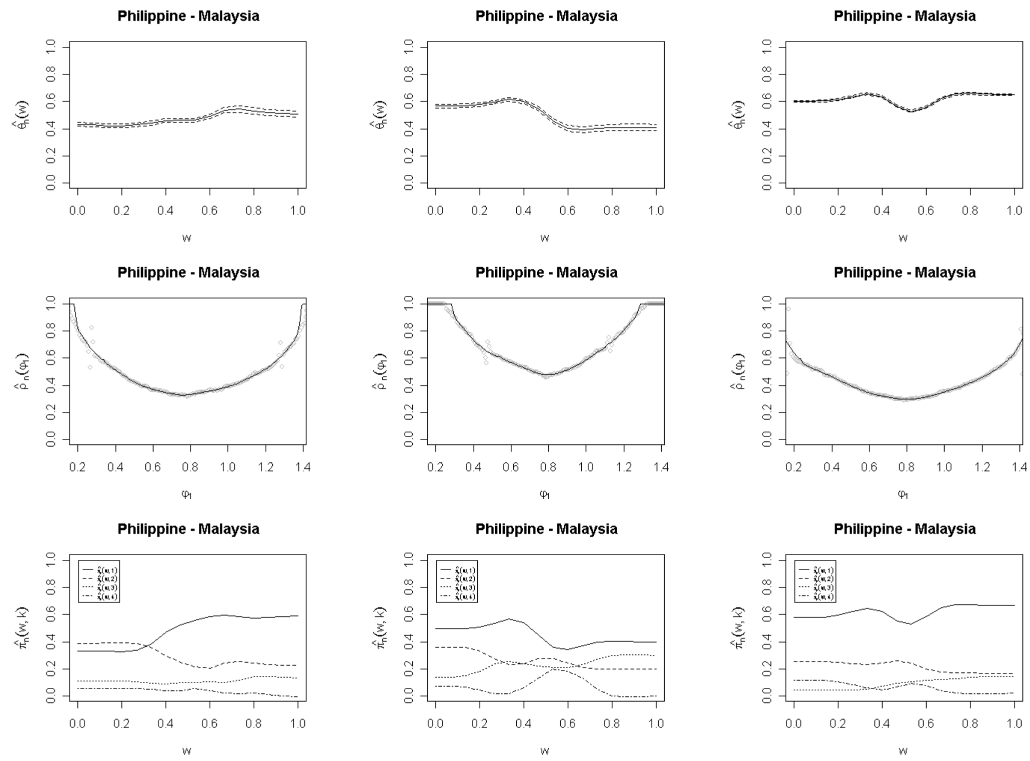

FIGURE 3.B.5. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Philippine vs Malaysia.

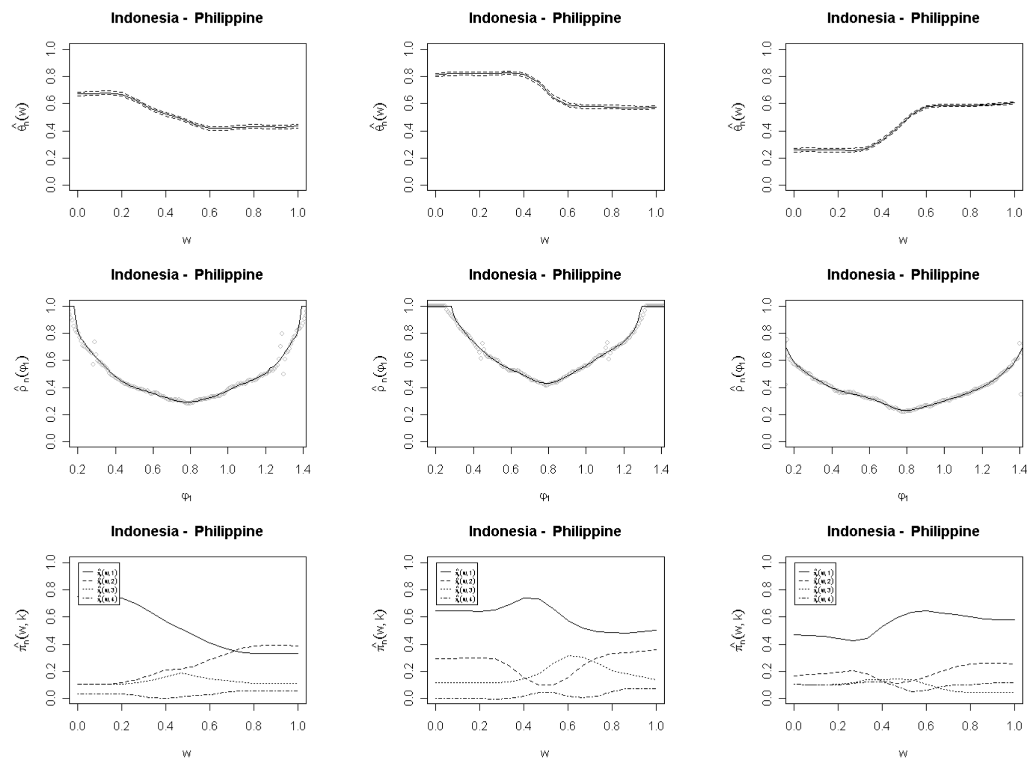

FIGURE 3.B.6. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Philippine vs Indonesia.

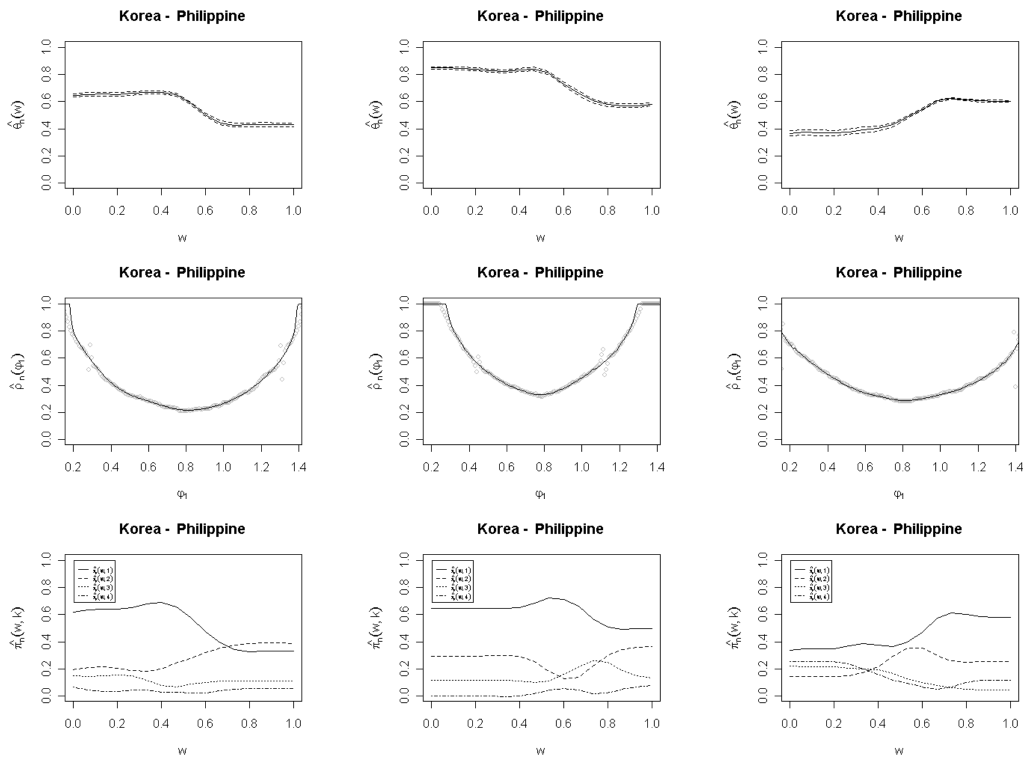

FIGURE 3.B.7. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Philippine vs Korea.

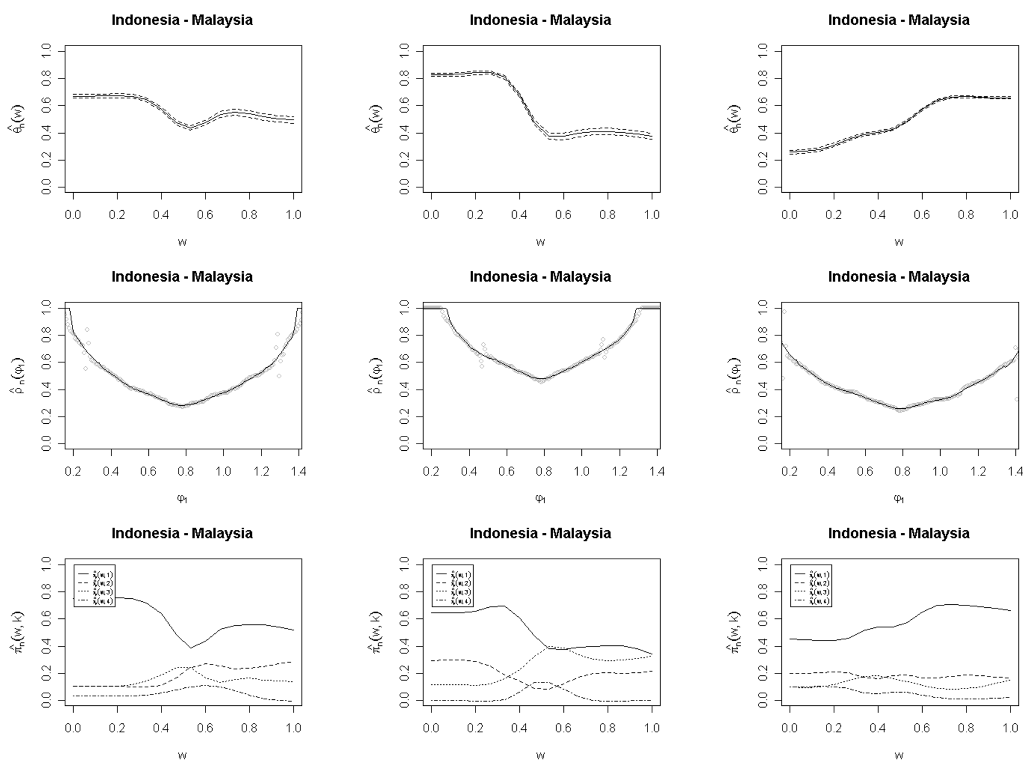

FIGURE 3.B.8. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Indonesia vs Malaysia.

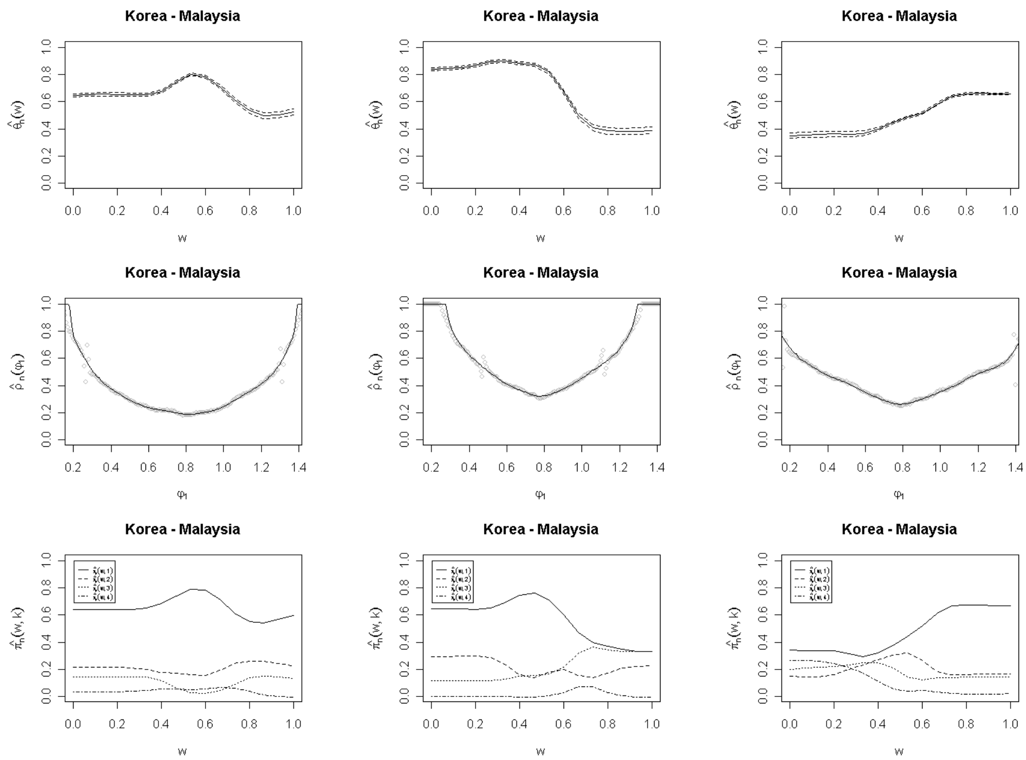

FIGURE 3.B.9. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Malaysia vs Korea.

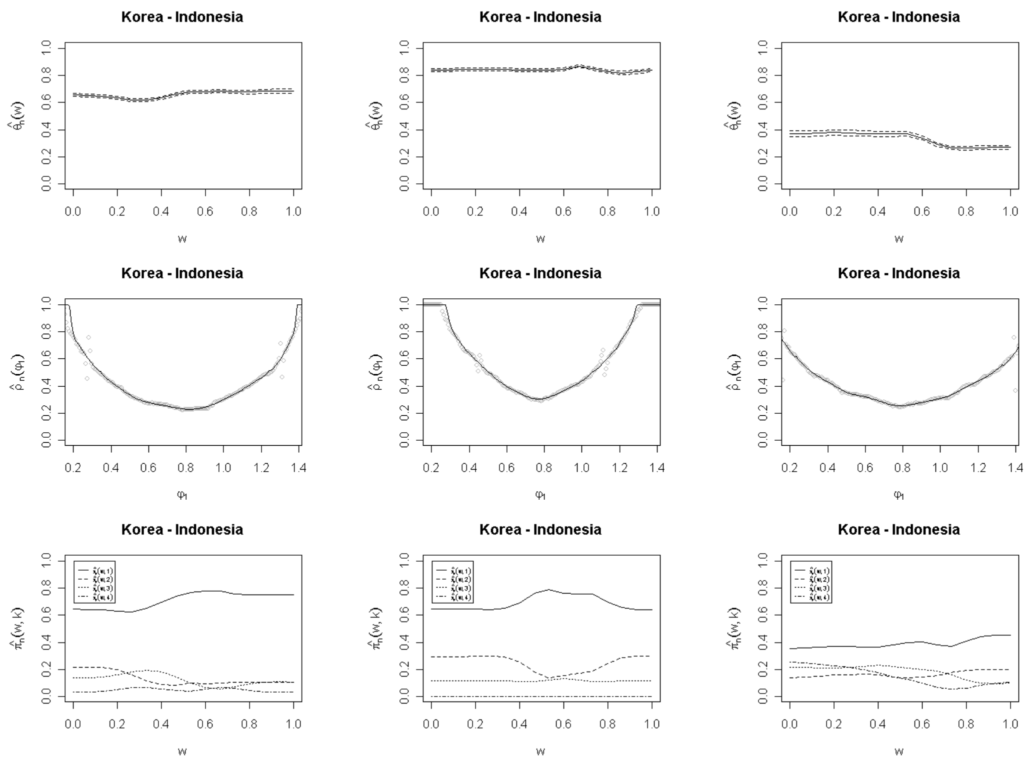

FIGURE 3.B.10. Results of the bivariate extremal index, bivariate cluster size probability and tail dependence for Indonesia vs Korea.

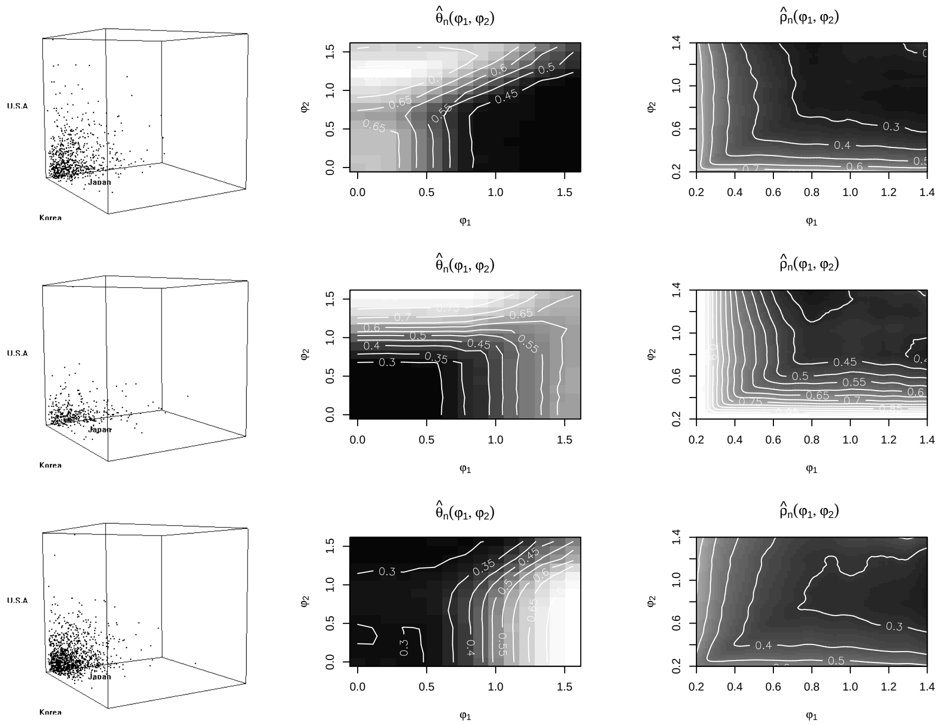

FIGURE 3.B.11. Results for Korea, U.S.A and Japan returns (left), the trivariate extremal index (middle) and the trivariate tail dependence function (right), for the pre, the post and the period crisis (top, middle and bottom panel respectively).

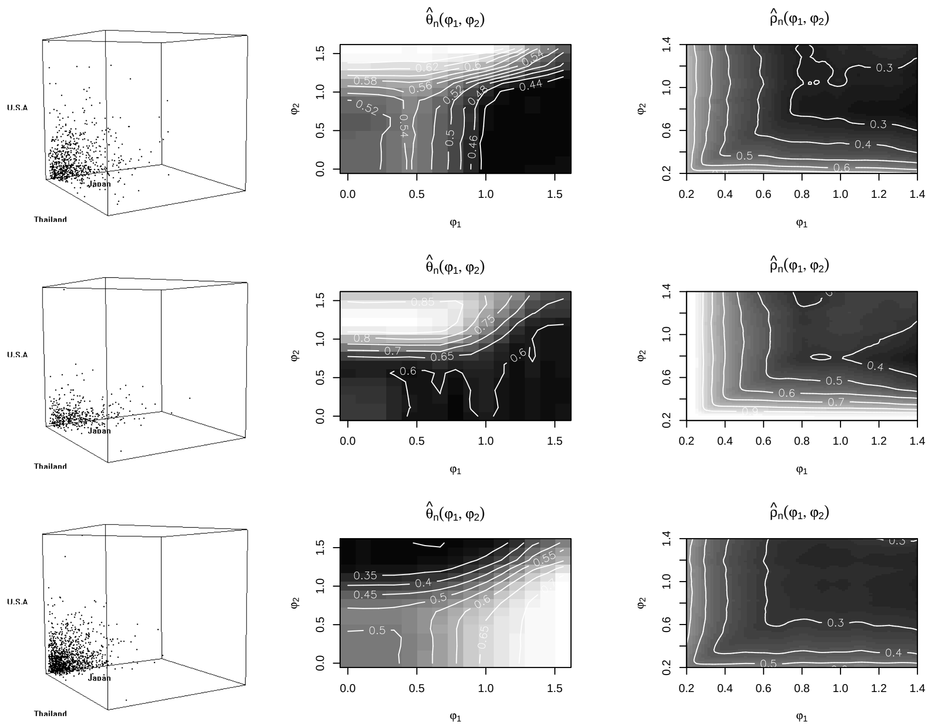

FIGURE 3.B.12. Results for Thailand, U.S.A and Japan returns (left), the trivariate extremal index (middle) and the trivariate tail dependence function (right), for the pre, the post and the period crisis (top, middle and bottom panel respectively).

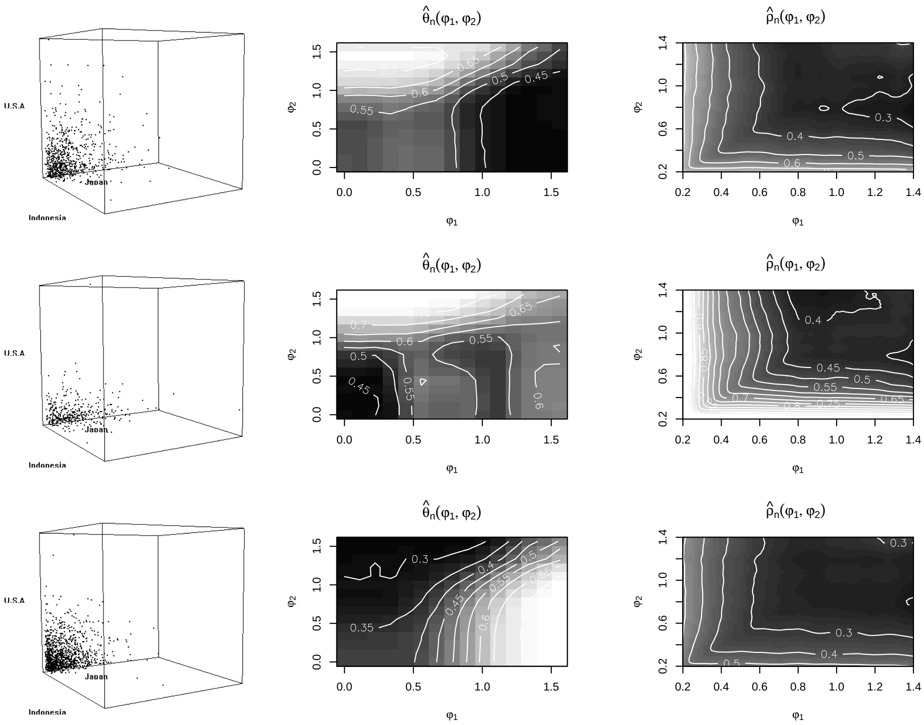

FIGURE 3.B.13. Results for Indonesia, U.S.A and Japan returns (left), the trivariate extremal index (middle) and the trivariate tail dependence function (right), for the pre, the post and the period crisis (top, middle and bottom panel respectively).

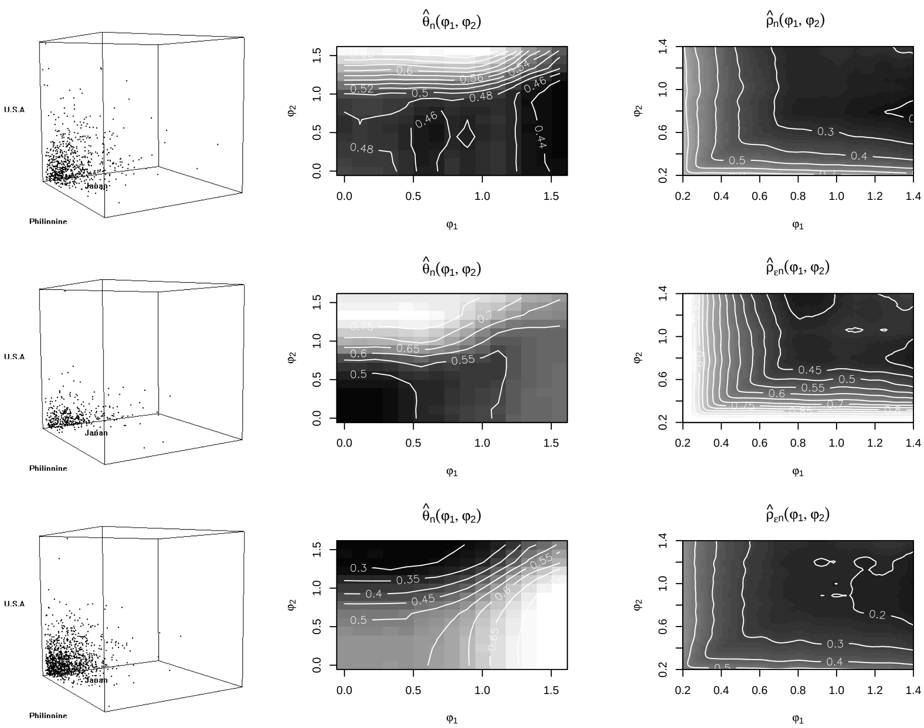

FIGURE 3.B.14. Results for Philippine, U.S.A and Japan returns (left), the trivariate extremal index (middle) and the trivariate tail dependence function (right), for the pre, the post and the period crisis (top, middle and bottom panel respectively).

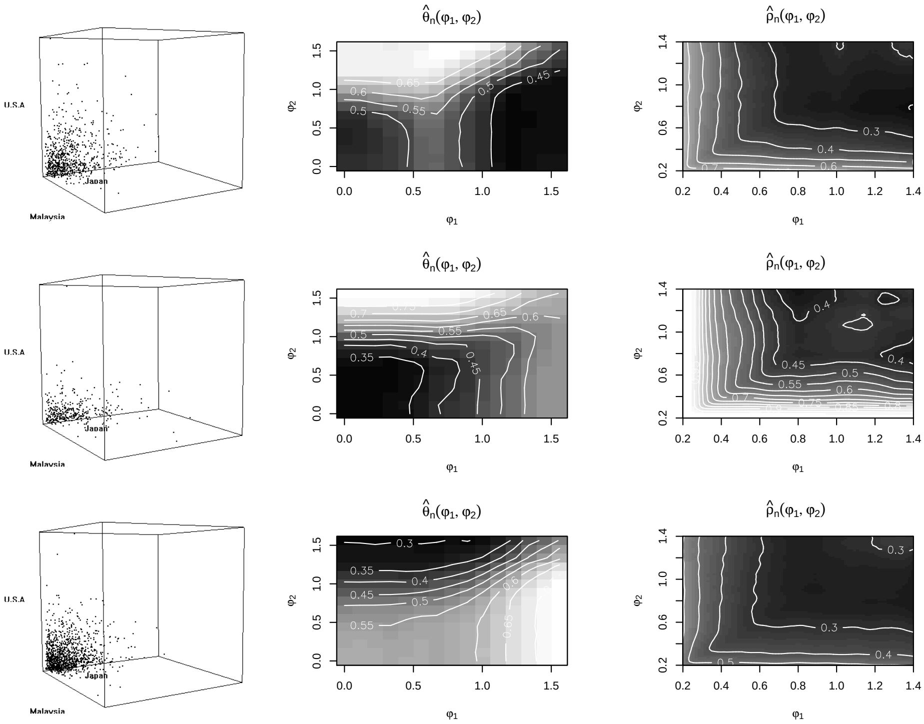

FIGURE 3.B.15. Results for Malaysia, U.S.A and Japan returns (left), the trivariate extremal index (middle) and the trivariate tail dependence function (right), for the pre, the post and the period crisis (top, middle and bottom panel respectively).

CHAPTER 4

Multivariate threshold methods with self-exciting functions

"The financial markets generally are unpredictable.
So that one has to have different scenarios.
The idea that you can actually predict what's going to happen
contradicts my way of looking at the marke."
(George Soros)

4.1. Introduction

A key problem in financial econometrics is modelling, estimation and forecasting of the dependence structure of financial assets returns. A great deal is known about the second moments of financial asset returns, sometimes associated to the cause of cluster in the extremes. Popular parametric models for volatility include the ARCH-GARCH family (Engle (1982b)) and the stochastic volatility (SV) family (Clark (1973)). In these models volatility is usually extracted from daily squared returns, which are unbiased but noisy estimates of daily conditional volatility.

More generally, the study of extreme dependence may reveal contrasts which are obscured when we only concentrate on examining the conditional second moment. For instance, Jalal and Rockinger (2004) investigated the consequences for value at risk and expected short fall purposes of using a GARCH filter on various misspecified processes. They show that careful investigation of the adequacy of the GARCH filter is necessary since under misspecifications a GARCH filter appears to do more harm than good. Poon et al. (2003) using a range of volatility filters find that tail index and tail dependence can be partially captured by models for heteroscedasticity. Moreover, they find that there is no clear reason to prefer one volatility filter over another. Also they note that tail index estimates are significantly reduced when returns are filtered for heteroskedasticity and that the reduction is the most dramatic when the SV model is used.

In addition to this Davis and Mikosch (2006b) showed that unlike the situation for GARCH processes (Davis and Mikosch (2006a)), there is no extremal clustering for SV processes in both the light- and heavy-tailed cases. That is, large values of the processes do not come in clusters. More precisely, the large sample behaviour of maxima is the same as that of the maxima of the associated iid sequence. So while both stochastic volatility and GARCH processes exhibit volatility clustering, only the GARCH has clustering of extremes.

While these models imply some information about extreme events still little is known about the extremes per se. For this reason, it is an advantage to have techniques that are focused purely on extreme movements, and are not influenced by the degree of temporal dependence in more routine circumstances. The resulting information about the degree to which extreme losses are more likely following earlier extreme losses is particularly valuable in the measurement of risk over fixed time intervals.

Several authors such as Embrechts et al. (1997); Matthys and Beirlant (2001); McNeil (1998); McNeil and Frey (2000); McNeil et al. (2005) and Novak (2004) have argued that extreme value theory allows to take into account explicitly the rare events contained in the tails. This presents three main advantages over classical methods such as conditional models (GARCH or SV processes), historical simulations or normal distribution approach.

Extreme value methodology focuses on modelling the tail behaviour of a distribution using only extreme values rather than the whole data set. Furthermore, as we do not assume a particular model for returns, we allow the data to speak for itself in order to fit the distribution tails.

The standard Peaks over threshold (POT) model in extreme value theory, which describes the appearance of extremes in iid data, elegantly subsumes the models for maxima and the Generalized Pareto distribution (GPD) models for excess losses. However, the characteristics of financial return series such as clustered extremes and serial dependence typically violate the assumptions of independence in the POT model. These problems are often addressed by the application of a declustering method, and then the standard model is fitted to cluster maxima only.

Another form to deal with the cluster on extremes is to use a marked self-exciting version of POT model introduced preliminarily in McNeil et al. (2005); Chavez-Demoulin et al. (2005); Herrera and Schipp (2009), where the cluster of extreme data will be modelled as self-exciting point processes without involving a prefiltration of data. The main characteristic of these models is that the intensity of occurrence of extreme events can depend on past extreme events and the size marks (predictable) allowing more realistic models.

Point Process of this kind has proved to be an efficient tool to model earthquake occurrences. Models as of Hawkes the ETAS (Epidemic Type Aftershock Sequence) or Stress Release models are considered to be standard branching model.

In the multivariate case the classical extreme value approaches for iid random variables are based on the Pickands dependence function, the estimation of an exponent measure or a Copula function.

In this chapter, we generalize the results on self-exciting point process approaches to the multivariate in a similar form as in Smith et al. (1997). In particular, the likelihood for this multivariate threshold likelihood model is based on properties of the exponent measure of a multivariate extreme distribution and the marginal observations below their respective threshold as censored at the threshold.

This chapter makes several contributions to the empirical literature on extreme dependence. An innovating feature of the present chapter, is that, contrary to earlier works on multivariate extreme models, the temporal dependence structure of each underlying marginal is not treated with an ARCH, GARCH, Multifractal Random Walk, etc., filter, which certainly impacts on the dependence properties among the marginals, i.e., the Copula function (or the Pickand's dependence function or the exponent measure) by the fact that they concentrate principally on the center of the distribution (see Malevergne and Sornette (2005) Section 5.3.4 for comments in this issue). Instead of forcing a single distribution for the entire sample, it is possible to investigate only the tails of the sample distribution using the models proposed, if only the tails are important for practical purposes.

Taking into account only the extreme events in the model yields to an important improvement in the quality of the multivariate dependence estimation because it may influence profoundly results on multivariate extreme events.

Furthermore, the second contribution in this chapter in the is the concept of dynamical dependence among marginals, which has been recently introduced in risk management by Dias and Embrechts (2004), Patton (2006) and Giacomini et al. (2007) for instance. In all these papers, the authors investigate the dynamical evolution of the multivariate dependence through the existence of structural changes in the dependence.

We follow these authors, however, under the point of view of the exponent measure and not of copulas, for the multivariate extreme dependence. We propose two parametric models to take into account the multivariate extreme dependence and a semiparametric model in the spirit of the model Dias and Embrechts (2004), but with other structure to determine the time of change of the multivariate dependence, which has the characteristic of time-invariant.

The third important contribution of this chapter is the application of these models to the three most important stock market indices from U.S future markets. Different examples of stress testing scenarios were calculated to the most important extreme events during the period of study, as the mini crash of 1997, the Russian default of 1998, the Dot.com crash of 2000, and the extreme movements by the attacks on September 11, 2001.

This chapter is organized as follows. In Section 4.2 we outline relevant aspects of the classical POT model, and then in Section 4.3 we describe the Self-exciting POT model theory that are central to the chapter and discusses a conditional intensity based approach to inference for point processes. In section 4.4 we develop different approaches to the multivariate case, allowing even the dependence to be time varying. Section 4.5 illustrates some applications to the most important extreme events in the last years in U.S stock markets. Conclusions and future works are resumed in Section 4.6.

4.2. The classical procedure

In this section we summarize the main results from POT method which underlie our modelling. General literature on the subject of extreme value theory includes Embrechts et al. (1997); McNeil et al. (2005); Falk et al. (2004).

Let $\mathbf{X}_n = \{X_{1n}, \ldots, X_{dn}\}$ be a sequence of iid multivariate random variables with distribution function $F(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ and let

$$\mathbf{M}_n = \left\{ \bigvee_{i=1}^{n} X_{i1}, \ldots \bigvee_{i=1}^{n} X_{id} \right\}$$

be the vector of componentwise maxima of $\mathbf{X}_n$ with distribution function $F^n(\mathbf{x})$. Assume that there exist sequences of normalizing d- variate sequences $\mathbf{a}_n$ and $\mathbf{b}_n \in \mathbb{R}^d$, and distribution function $G(\mathbf{x})$ with nondegenerate marginals such that

$$\lim_{n\to\infty} F(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) = G(\mathbf{x}) \Longleftrightarrow \lim_{n\to\infty} n\{1 - F(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n)\} = -\log G(\mathbf{x}) \qquad (4.2.1)$$

for all $\mathbf{x}$ such that $G(\mathbf{x}) > 0$.

From (4.2.1), all the marginal distributions of a Multivariate extreme value distribution (MVE) are, if not degenerated, again MVED of lower dimensions and univariate marginals are Generalized extreme value distributions (GEV). In particular, if F_i are each marginal of F being in the Maximum domain of attraction of G_i, $F_i \in MDA\left(G_i\right)$, where G_i are the corresponding component of G then, by the Fisher-Tippett theorem, G_i belongs to the type of the distribution

$$H_{\xi}\left(x\right)=\begin{cases}\exp\left\{-\left(1-\xi x\right)^{-1/\xi}\right\}, & 1+\xi x>0,\ \ \xi\neq 0\\ \exp\left\{-e^{-x}\right\}, & -\infty<x<\infty,\ \ \xi=0.\end{cases}$$

H_{ξ} is known as the Generalized extreme value distribution. For $\alpha>0$, $\Phi_{\alpha}:=H_{\alpha^{-1}}\left(\alpha\left(x-1\right)\right)$ is the standard Frèchet distribution, $\psi_{\alpha}\left(x\right):=H_{-\alpha^{-1}}\left(\alpha\left(x+1\right)\right)$ is the standard Weibull distribution, and $\Lambda\left(x\right):=H_0\left(x\right)$ is the standard Gumbel distribution.

Now from the well known Pickands-Balkema-de Haan Theorem, the excess distribution function $F_u(x)=P\left(X-u\leq x \mid X>u\right),\ x\geq 0$ for a high threshold $x\geq u$ of all marginals $F_i\in MDA(H_{\xi})$ can be well approximated by a Generalized Pareto distribution (GPD), that is

$$\lim_{u\uparrow x}\ \sup_{0<x<x_F-u}\left|F_u(x)-G_{\xi,\beta}(x)\right|=0,$$

where x_F is the right endpoint and $G_{\xi,\beta}$, with parameters $\xi\in\mathbb{R}$ and $\beta>0$, has distribution tail

$$\bar{G}_{\xi,\beta}(x)=\begin{cases}\left(1+\xi\frac{x}{\beta}\right) & if\ \xi\neq 0,\\ \exp(-x/\beta) & if\ \xi=0,\end{cases}$$

and $x\in D(\xi,\beta)$

$$D(\xi,\beta)=\begin{matrix}[0,\infty) & if\ \xi\geq 0,\\ [0,-\beta/\xi] & if\ \xi<0.\end{matrix}$$

From a different point of view Smith (1989) introduces, via a point process representation of the exceedances, an elegant approach to the Peaks over a threshold $x\geq u$ (POT) for the univariate case.

Until now we have only considered the magnitude of excess losses over high thresholds the point process approach consider exceedances of thresholds as events occurring in time and use a point process approach to model these events. The idea is to view all events exceeding a given level x as a bidimensional point process with state space $\chi=(0,1]\times(u,\infty)$. The intensity of this process is defined for all Borel sets if we can define it on all rectangle of the form $A=(t_1,t_2)\times(x,\infty)$ where t_1 and t_2 are time coordinates and $x\geq u$ is a given threshold level of the process. The point process is the defined by

$$N(A)=\sum_{i=1}^{n}\mathbb{I}_{\{(i/n,X_i)\in A\}}.$$

If the process is stationary and satisfies a condition that there are asymptotically no clusters among at extremes, then its limiting form is a non-homogeneous Poisson process and the intensity at a point (t,x) is given by

$$\lambda(t,x)=\frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)_{+}^{-1/\xi-1}, \qquad (4.2.2)$$

where $y_+ = \max(x, 0)$ and μ,σ,ξ represent respectively a location parameter, scale parameter and shape parameter. Note that this does not depend on t hence the two-dimensional Poisson process is non-homogeneous. Therefore, the intensity measure of the set A for any $x \geq u$ may be expressed in the form of an one-dimensional homogeneous Poisson process $\lambda(x) = \lambda(x, t)$ with rate $\tau(x) = -\ln H_{\xi,\mu,\beta}$ as

$$\Lambda(A) = \int_{t_1}^{t_2} \int_{x}^{\infty} \lambda(t, y) d_y dt = -(t_2 - t_1) \ln H_{\xi,\mu,\sigma}.$$

Now the tail of the excess over the threshold u, denoted $\overline{F}_u(x)$, can be calculated as the ratio of the rates of exceeding the levels $u + x$ and u as follows

$$\overline{F}_u(x) = \frac{\overline{F}(x+u)}{\overline{F}(u)} = \frac{\tau(u+x)}{\tau(u)} = \left(1 + \frac{\xi x}{\sigma + \xi(u-\mu)}\right)^{-1/\xi} = \overline{G}_{\xi,\beta}(x),$$

where $\beta = \sigma + \xi(u - \mu)$ is simply a scaling parameter.

A useful reparametrization of the equation (4.2.2) for the next section is rewritten this in terms of the scale parameter $\beta = \sigma + \xi(u - \mu)$ and the rate of the one-dimensional Poisson process of exceedances of the level u, $\tau(u) = \tau = -\ln H_{\xi,\mu,\beta}(u)$

$$\lambda(x) := \lambda(t, x) = \frac{\tau}{\beta}\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi-1}, \tag{4.2.3}$$

where $\xi \in \mathbb{R}$ and $\tau, \beta > 0$.

The aim of presenting the two dimensional Point process derivation of the POT method is for introducing easily the new methods, which have as basis the point process representation.

Returning to the multivariate case there are several different but equivalent characterizations of multivariate extreme value distributions, see for examples the books of Resnick (1987), Falk et al. (2004) and Galambos (1978). One feature of multivariate extreme value distributions is that, the dependence structure is preserved under transformations of the marginal distributions, so there is no loss of generality in restricting attention to a particular univariate extreme value family. For this reason we assumed unit Frèchet marginals. In this chapter we consider the next definition of MEV distribution.

Definition 4.2.1. A multivariate distribution $G(x)$ is a MEV distribution with unit Frèchet marginals if, and only if, its joint distribution $G : \mathbb{R}^d \to [0, 1]$ can be expressed as

$$\begin{aligned} G(x) &= \exp\left\{-\int_{\mathbb{S}^{d-1}} \bigvee_{j=1}^{d} \left(\frac{w_j}{x_j}\right) \mathcal{H}(d\mathbf{w})\right\} \\ &= \exp\left\{-\mu\left([0, \mathbf{x}]^c; \theta\right)\right\} \end{aligned}$$

subject to the condition

$$\int_{\mathbb{S}^{d-1}} w_j \mathcal{H}(d\mathbf{w}) = 1, \qquad j = 1, \ldots, d$$

where $\mathbb{S}^{d-1} = \left\{(w_1, \ldots, w_{d-1}) : \sum_{j=1}^{d-1} w_j \leq 1,\ w_j \geq 0,\ j = 1, \ldots, d-1\right\}$ is the $(d-1)$ dimensional unit simplex in $\mathbb{R}^d$, $\mathcal{H}$ is a positive finite measure on $\mathbb{S}^{d-1}$ and $\mu\left([0, \mathbf{x}]^c; \theta\right)$ is referred as the *exponent measure function* and θ is a collection of parameter that characterizes the structure of the dependence of the exponent measure.

Contrary to the univariate case, there is no finite parametrization for the exponent measure function $\mu\left([0,\mathbf{x}]^c;\theta\right)$. Thus, it is common to use specific parametric families for $\mu\left([0,\mathbf{x}]^c;\theta\right)$ such as the logistic (Galambos (1978)), the asymmetric logistic (Tawn (1988a)), the negative logistic (Galambos (1975)) or the Polynomial models (Klüppelberg and Mayer (2006)).

4.3. Self-exciting extreme value models

The problem with the theory outlined in the previous section is that it assumes that the underlying series is independent, which is unrealistic in most of applications. Serial dependence and volatility cluster play an important role in the most applications on returns of financial series, and so exceedances of a high threshold for daily financial return series do not necessarily occur according to a homogeneous Poisson process. Therefore, the direct application of the POT method is nonviable.

However, under weak conditions the POT representation may be applied to the maximum value of each cluster. The problem here is the identification of independent clusters of exceedances over a high threshold. This is because of the fact that the choice of declustering scheme has often an important impact on estimates of cluster characteristics.

Possible algorithms are given by the run method, the block method or the interval method (see Beirlant et al. (2004)). In particular the interval estimator introduced by Ferro et. al. Ferro and Segers (2003) propose an automatic declustering scheme that is justified by an asymptotic result for the arrival times between threshold exceedances. The scheme relies on the estimation of *extremal index* prior to declustering, which can be interpreted as the reciprocal of the mean cluster size. However, this method consists of a two steps procedure and the cluster behaviour of the extremes is lost.

By other way, the methodology introduced in this section takes advantage of the structure of the model, thus allowing the existing (dependent) data to be used more effectively.

In the early 1970s, Hawkes (1971) introduced a family of what he called "self-exciting" or "mutually exciting" models, which became both pioneering examples of the conditional intensity methodology. The models have been greatly improved and extended by Ogata (1988), whose ETAS model has been successfully used to elucidate the detailed structure of aftershock sequences.

In these models a recent spate of threshold exceedances causes the instantaneous risk of a threshold exceedances at a particular time to be higher. As McNeil et al. (2005) suggest, the structure of these processes, which has traditionally been used in the modelling of earthquakes, would also seem appropriate for modelling market shocks and the tremors that follow these.

In the next section, we bring together some of the most widely used concepts of point process models. We extend the simple Poisson process in the same way that the cluster processes did. Furthermore, we introduce the main concepts in marked point process. These processes could be already covered formally by the general theory of point processes, as they can be represented as a special type of point process on a product space. However, marked point processes are worth additional studying because of their wide range of applications and their conceptual importance.

The results shown in the next section are based on first volume of Daley and Vere-Jones (2003).

4.3.1. Preliminaries. Unless stated otherwise, all random elements occurring in this paper are supposed to be defined within a complete, separable metric space (c.s.m.s) $\mathcal{T}$, and will generally interpret as $\mathbb{R}$. A point process is defined as a random counting measure on $\mathcal{T}$, meaning technically a measurable mapping from a probability space $(\Omega, \mathcal{E}, \mathcal{P})$ into the space $\left(\mathcal{N}_{\mathcal{T}}^{\#}, \mathcal{B}\left(\mathcal{N}_{\mathcal{T}}^{\#}\right)\right)$.[1] Furthermore, let $\{\mathcal{H}_t,\ t \geq 0\}$ be a non-decreasing family of right continuous σ-algebras $\mathcal{H}_s \subset \mathcal{H}_t \subset \mathcal{H}$ for any $0 \leq s \leq t$.

Definition 4.3.1. A simple point process is a sequence of random variables $\{t_i\}_{i \in \{1,\ldots,n\}}$ satisfying.

(1) $\mathbb{P}\left(0 < t_1 \leq \cdots \leq t_n\right) = 1$,
(2) $\mathbb{P}\left(t_n \leq t_{n+1},\ t_n < \infty\right) = \mathbb{P}\left(t_n < \infty\right) \quad (n \geq 1)$,
(3) $\mathbb{P}\left(\lim_{n\to\infty} t_n = \infty\right) = 1$.

To describe a point process it is sufficient to specify its intensity function. Then, let $N(t) := \sum_{i\geq 1} \mathbb{I}_{\{t_i \leq t\}}$ be the right-continuous counting function. Define $\lambda(t)$ as the intensity of $N(t)$, that is a positive $\mathcal{H}_t$ predictable process , i.e, $\lambda(t)$ is adapted to $\mathcal{H}_t$, and left-continuous with right hand limits, with unique compensator

$$\Lambda(t) = \int_0^t \lambda(u)\, du,$$

for $N(t)$.

Assuming that it exists, the conditional intensity $\lambda(t \mid \mathcal{H}_t)$ is defined as the unique non decreasing, $\mathcal{H}$-predictable process

$$\lambda(t \mid \mathcal{H}_t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbb{E}\left(N[0, t+\delta t) - N[0,t) \mid \mathcal{H}_t\right).$$

Example 4.3.2. Suppose that $t_1, t_2 - t_1, t_3 - t_2, \ldots$ are independent and distributed according to an exponential distribution with parameter λ. Then $N(t) := \sum_{i\geq 1} \mathbb{I}_{\{t_i \leq t\}}$ is called a homogeneous Poisson process with intensity λ.

Given a univariate point process $\{t_n\}$, one often has additional information on the points t_n. It is modelled by a random element X_i which is called the mark of t_n. In our case the marks will be referred to the exceedances over a threshold for a stationary sequence.

Definition 4.3.3. A marked point process (MPP), with locations in the c.s.m.s $\mathcal{T}$ and the marks in the c.s.m.s $\mathcal{X}$, is a point process $\{(t_i, X_i)\}$ on $\mathcal{T} \times \mathcal{X}$ with the additional property that the ground process $N_g(t)$, the process of the locations $\{t_n\}$ where the events occur, is itself a point process; i.e, for bounded $A \in \mathcal{B}_{\mathcal{T}}$, $N_g(A) = N(A, \mathcal{X}) < \infty$.

[1] $\mathcal{N}_{\mathcal{T}}^{\#}$ is the metric space of all boundedly finite counting measures on $\mathcal{T}$ and $\mathcal{B}\left(\mathcal{N}_{\mathcal{T}}^{\#}\right)$ is a family of Borel sets that can be used to define measures on $\mathcal{N}_{\mathcal{T}}^{\#}$.

In general, a MPP N is defined as a set of pairs $\{(t_i, x_i)\}$ on the product space $\mathcal{T} \times \mathcal{X}$, with the the ground process N_g being the marginal process of locations, or as ordinary point process in $\mathcal{T}$, $\{t_i\}$ say, with an associated sequence the random variables (marks) $\{x_i\}$ taking their values in $\mathcal{X}$.

There exists a rich class of MPPs with a great variety of forms that can be taken by the marks and the variety of dependence relations among marks and locations. Two important classes are defined below

Definition 4.3.4. .

(1) The MPP N is simple if the ground process N_g is simple.
(2) The MPP N on $\mathcal{T} = \mathbb{R}$ is stationary (homogeneous) if the probability structure of the process is invariant under shifts in $\mathcal{T}$.

The next definition characterizes two important types of independence relating to the mark structure of MPPs.

Definition 4.3.5. Define a MPP $N = \{(t_i, x_i)\}$ on $\mathcal{T} \times \mathcal{X}$ to have

(1) independent marks, if given the ground process N_g the marks $\{x_i\}$ are mutually independent random variables such that the distribution of the x_i, in our case the exceedances, depends only on the corresponding location t_i.
(2) unpredictable marks for $\mathcal{T} = \mathbb{R}$, if the distribution of the mark at t_i is independent of the locations and marks $\{(t_j, x_j)\}$ for which $t_j < t_i$.

Of course the most common case of an MPP with independent marks occurs when the marks are iid. Similarly, the most common case of a process with unpredictable marks occurs when the marks are conditionally iid given the past of the process. The most interesting extensions appear when we drop the assumption of completely independent marks and consider ways in which either the marks can influence the future development of the process or the current state of the process can influence the distribution of marks, or both. Such class of processes are the marked self-exciting processes.

Definition 4.3.6. A linear marked self -exciting process, is defined as the process which intensity has the form

$$\lambda(t \mid \mathcal{H}_t) = \mu(t) + \phi \sum_{i:t_i<t} g(t - t_i, x) \tag{4.3.1}$$

$$= \mu(t) + \phi \int_0^t g(t - ds, x) N_x(ds \times dz) \tag{4.3.2}$$

here $\{\mu(t)\}_{t\in\mathbb{R}}$ is a given nonnegative stochastic process which is locally integrable, $\phi \geq 0$ and $g(\cdot)$ denotes a non-negative measurable function, and $\int_0^t g(u) dN(u)$ is the stochastic Stieljes integral of the process g with respect to the counting process N_g.

One of the most important characteristics of these processes is that they combine in the model both a cluster process representation and a simple conditional intensity representation, which is moreover linear. In equation (4.3.1) μ can be interpreted as an exogenous rate, whereas g can be interpreted as an endogenous rate and that explains why these processes can be useful in many applications in finance.

The particular feature of the self-exciting model is the representation of the conditional intensity as a sum of contributions from all previous events. This process can also be interpreted as a "branching process with immigration", the immigration component being described by the constant rate (in time) Poisson process of background events, and the "offspring" from a given "immigrant". The events immigrants arrive according to a Poisson process at the constant rate μ, while the offspring arise as elements of a finite Poisson process that is associated with some point already constructed.

An important task is to find conditions that ensure the existence of stationary in this process, i.e., of realizations of point sets $\{t_i\}$ on the whole space $\mathcal{T} = \mathbb{R}$ having structure above and with distribution invariant under translation. The next proposition states this formally.

PROPOSITION 4.3.7. *Sufficient conditions for the existence of a stationary marked Hawkes process are*

(1) *the intensity measure* $\mu(\cdot)$ *is totally finite,*
(2) $\mathbb{E}[\psi(x)] = \int_{\mathcal{X}} \psi(x) F(dx) < 1$,*where the factor* $\psi(x)$ *determines the relative average sizes of families with different marks and* $F(dx)$ *is the mark distribution.*

In this study we use four types of self-exciting functions for the MPP. Before we introduce these functions, we take a look how these will be used in the Peaks over threshold method.

4.3.2. Peaks over threshold model as a marked point process with self-exciting function. Given a sequence of events $t_1, \ldots, t_n$ we define a point process of exceedances $N(\cdot) = \sum_{i=1}^{n} I_{\{(t_i, \tilde{x}_i) \in \cdot\}}$ with state space $(0, n] \times (u, \infty)$ for $k = 1, \ldots, N_u$, where t_k are the times of occurrence of the extreme vents, that is, the excedances over a high threshold u, $\tilde{x}_k$ are the marks, in our case the magnitude of the excesses for $x_i > u$, i.e., $\tilde{x_i} = (x_i - u) \vee 0$, and N_u are the number of exceedances.

Following the approach outlined in Section 4.2 we modify the equation (4.2.3) by incorporating a function to model the dependence of the frequencies and sizes of the events over a high threshold u.

We begin replacing the rate of the one-dimensional Poisson process of exceedances of the level u, τ and the scale parameter β in (4.2.3) by their self-excitement versions

$$\tau(t) = \tau + \phi w(t) \tag{4.3.3}$$

and

$$\beta(t) = \beta + \eta w(t), \tag{4.3.4}$$

where $w(t) = \sum_{k:t_k<t} g(t - t_k;\, x_k)$ is defined as in (4.3.1) and $\eta, \phi > 0$.

Note that the only difference that we are adopting is τ and β depending on both marks and time points of all preceding events at times up to but not including t according to a common self-exciting function.

It implicates that the exceedances over the threshold u occur according to a one dimensional self-exciting marked point process with GPD density and conditional intensity described by the equation

$$\lambda^*(t,x) = \frac{(\tau + \phi w(t))}{\beta + \eta w(t)} \left(1 + \xi \frac{x-u}{\beta + \eta w(t)}\right)^{-1/\xi - 1}. \tag{4.3.5}$$

Furthermore, the conditional rate of crossing the threshold u at time t is defined as follows

$$\tau(t,x) = \int_x^\infty \lambda^*(t,y)dy = (\tau + \phi w(t)) \left(1 + \xi \frac{x-u}{\beta + \eta w(t)}\right)^{-1/\xi}.$$

Therefore, the implied distribution of the excess losses when an exceedance takes place is generalized Pareto

$$\begin{aligned} P\left(X_k > u + x \mid T_k = t, \mathcal{H}_t\right) &= \frac{\tau(t, x+u \mid \mathcal{H}_t)}{\tau(t, x \mid \mathcal{H}_t)} \\ &= \bar{G}_{\xi, \beta + \eta w(t)}(x). \end{aligned}$$

In the next proposition we resume the main properties of this model.

PROPOSITION 4.3.8. *Let X be stationary sequence, whose exceedances over threshold u occur according to an one dimensional self-exciting marked point process with GPD density and conditional intensity as it is described by the equation (4.4.1) then,*

(1) *If $\phi = \eta = 0$ we have the traditional POT model described in (4.2.3).*
(2) *If $\eta = 0$ we have a MPP with self-exciting function where the marks have GPD iid, that is with unpredictable marks.*
(3) *If $\eta > 0$ we have a MPP with self-excitement function where the marks are conditionally GPD, that is with predictable marks.*
(4) *If $\phi > 0$ the estimated intensity of exceeding the threshold in the model is modelled by a self-exciting function.*

Now we define the class of self-exciting functions that we wish to use.

DEFINITION 4.3.9. Let N be a stationary MPP with Generalized Pareto densities for the marks defined as the exceedances $x \geq u$, where u is a high threshold. Then, the self-exciting function $g(\cdot, \cdot)$ can take the following forms

(1) The generalized Hawkes of order K

$$g(t,x) = \sum_{k}^{K} \phi_k (t-s)^{k-1} \exp(-\gamma t)(1 + \delta x), \text{ for } \delta, \gamma > 0. \tag{4.3.6}$$

(2) The simple Hawkes model

$$g(t,x) = \exp(\delta x - \gamma t), \text{ for } \delta, \gamma > 0. \tag{4.3.7}$$

(3) The Hawkes model with exponential decay

$$g(t,x) = \gamma (1 + \delta x) \exp(-\gamma t), \text{ for } \delta, \gamma > 0. \tag{4.3.8}$$

(4) The ETAS model introduced by Kagan and Knopoff (1987)

$$g(t,x) = (1+\delta x)\left(1+\frac{t}{\gamma}\right)^{-(1+\rho)}, \text{ for } \delta,\gamma,\rho>0. \tag{4.3.9}$$

In the first model we consider a generalization of the simple Hawkes model (see Hawkes (1971)), which is a generalized Poisson cluster process associating to cluster centers a branching process of descendants. Note that the third model is a Markovian model for $\delta = 1$ (see Daley and Vere-Jones (2003), pp. 243). The above specification can be extended to other classes of models as for example a marked stress release process.

4.3.3. Conditional intensities and likelihoods methods for marked point processes. In general, there are no easy methods for evaluating point process likelihoods on general spaces. However, the most important case for us $\mathcal{T} = \mathbb{R}$ can fortunately be resolved through conditional intensities.

An important definition for a finite point process distribution is the concept of Janosy measures.

DEFINITION 4.3.10. Define the points to be located in a c.s.m.s $\mathcal{X}$ with distribution $\{p_n\}$, which determines the total number of points in the population with $\sum_{n=0}^{\infty} p_n = 1$. Furthermore, for each integer $n \geq 1$, a probability distribution $\Pi_n(\cdot)$ is given on the Borel sets of $\mathcal{X}$ determining the joint distribution of the positions of the points of the process, given that their total number is n. Then, for any partition $(A_1,\ldots,A_n)$ of $\mathcal{X}$ the Janossy measure is defined as

$$J_n(A_1\times\cdots\times A_n) = p_n \sum_{\text{perm}} \Pi_n(A_{i_1},\ldots,A_{i_n})$$

where the summation $\sum_{\text{perm}}$ is taken over all $n!$ permutations $(i_1,\ldots,i_n)$ of the integers $(1,\ldots,n)$.

When the derivatives of the Janossy measure exist, then it takes a simple interpretation and it plays a fundamental role in the structural description and likelihood analysis of finite point process. For instance, if $j_n(x_1,\ldots,x_n)$ denotes the density of J_n in a c.s.m.s $\mathcal{X} = \mathbb{R}^d$ with respect to Lebesgue measure on $\mathbb{R}^d$ with $x_i \neq x_j$ for $i \neq j$. Then, $j_n(x_1,\ldots,x_n)\,dx_1\cdots dx_n$ is the probability that there are exactly n points in the process, one in each of the n distinct infinitesimal regions (x_i, x_i+dx_i).

We give now the definition of likelihood for finite point process in terms of the Janossy density.

DEFINITION 4.3.11. Define N to be a regular process, if for all integer $k \geq 1$ the local Janossy measures $J_k(dt_1\times\cdots\times dt_k \mid A)$ are absolutely continuous on $A^{(k)}$ with respect to Lebesgue measure $\mathcal{X}^{(k)}$. Then, the likelihood of a realization $t_1,\ldots,t_n$ of a regular point process N on bounded Borel set $A \subseteq \mathbb{R}^d$ is the local Janossy density

$$L_A(t_1,\ldots,t_n) = j_n(t_1,\ldots,t_n \mid A).$$

In order to define the conditional Intensities and likelihoods for MPPs, we defined first the states space on $[0,\infty)\times\mathcal{X}$. Furthermore, we need to fix on a measure in the mark space

$(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ to serve as a reference measure $\ell_{\mathcal{X}}(\cdot)$ (Lebesgue measure on $\mathcal{X}$) in forming densities. Now we are able to give an important definition concerning the *regularity* of a MPP.

DEFINITION 4.3.12. A MPP on $\mathcal{T} = \mathbb{R}^d \times \mathcal{X}$ is regular on A for a bounded Borel set $A \in \mathcal{B}_{\mathcal{X}}$ if for all $n \geq 1$ there exists a well defined Janossy measure J_n and therefore a Janossy density $j_n(\cdot \mid A \times \mathcal{X})$ as follows

$$j_n(t_1, \ldots, t_n, x_1, \ldots, x_n \mid A \times \mathcal{X})\, dt_1 \cdots dt_n \ell_{\mathcal{X}}(dx_1) \cdots \ell_{\mathcal{X}}(dx_n)$$

with the interpretation; the probability there exist points around $(t_1, \ldots, t_n)$ with marks around $(x_1, \ldots, x_n)$.

The following equivalences characterize completely a MPP and give the basis for the more general concept the likelihood ratio.

PROPOSITION 4.3.13. *Define $N(\cdot)$ as an MPP on $\mathbb{R}^d \times \mathcal{X}$, ℓ the Lebesgue measure on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$ and $\ell_{\mathcal{X}}$ the reference measure on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. Moreover, let A be a bounded set in $\mathcal{B}_{\mathbb{R}^d}$. Then, the following conditions are equivalent.*

(1) $N(\cdot)$ is regular on A.
(2) The ground process $N_g(\cdot)$ is regular on A, and for each $n > 0$ the conditional distribution of the marks $(x_1, \ldots, x_n)$, with locations $(t_1, \ldots, t_n)$ within A, is absolutely continuous with respect to $\ell_{\mathcal{X}}^{(n)}$ with density $f_{A,n}(x_1, \ldots, x_n \mid t_1, \ldots, t_n)$.
(3) If $\Pi(\cdot)$ is a probability measure equivalent to $\ell_{\mathcal{X}}$ on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, then $N(\cdot)$ is absolutely continuous with respect to the compound Poisson process $N_0(\cdot)$ for which the ground process N_0^g has positive intensity λ on A and the marks are iid with common probability distribution Π.

The essence of the approach for the likelihood of a MPP is the use of a causal description of the process through successive conditioning. For ease of writing, we use $j_n(t_1, \ldots, t_n \mid w)$ for the local Janossy density on the interval $(0, w)$ and $J_0(w)$ for $J_0((0, k))$. Furthermore, we define the intervals $\tau_i = t_i - t_{i-1}$ for $i \geq 1$ and $t_0 = 0$, and the conditional survivor functions $S_k(w \mid t_1, \ldots, t_{k-1}) = \mathbb{P}(\tau_k \mid t_1, \ldots, t_{k-1})$ and observe that these can be represented recursively in terms of the Janossy functions through the equations

$$\begin{aligned} J_0(T) &= S_1(T), \\ j_1(t_1, x_1 \mid T) &= p_1(t_1, x_1) = p_1(t_1) f_1(x_1 \mid t_1) \qquad (0 < t_1 < T), \\ j_2(t_1, t_2, x_1, x_2 \mid T) &= p_1(t_1) f_1(x_1 \mid t_1) p_2(t_2 \mid (t_1, x_1)) f_2(x_2 \mid (t_1, x_1), t_2) \quad (0 < t_1 < t_2 < T) \end{aligned}$$

where the $p_i(\cdot)$ are the densities, suitably conditioned, for the locations in the ground process, and the $f_i(\cdot)$ refer to the densities, again suitably conditioned, for the marks.

PROPOSITION 4.3.14. *For a regular point process on $\mathcal{T} = \mathbb{R}^+$, there exists a uniquely determined family of conditional probability density functions $p_n(t \mid t_1, \ldots, t_{n-1})$ and associated survivor functions*

$$S_n(t \mid t_1, \ldots, t_{n-1}) = 1 - \int_{t_{n-1}}^{t} p_n(u \mid t_1, \ldots, t_{n-1})\, du \tag{4.3.10}$$

defined on $0 < t_1 < \cdots < t_{n-1} < t$ *such that each* $p_n(\cdot \mid t_1, \ldots, t_{n-1})$ *has support carried by the half-line* (t_{n-1}, ∞)*, and for all* $n \geq 1$ *and all finite intervals* $[0, T]$ *with* $T > 0$*. Then, the following equations specify uniquely the distribution function of regular MPP on* $\mathbb{R}^+$

$$\begin{aligned}
J_0(T) &= S_1(T), \\
j_n(t_1, \ldots, t_n, x_1 \mid T) &\equiv j_n(t_1, \ldots, t_n, x_1 \mid (0, T)) \\
&= p_1(t_1) f_1(x_1 \mid t_1) p_2(t_2 \mid (t_1, x_1)) f_2(x_2 \mid (t_1, x_1), t_2) \cdots \\
&\quad \cdots p_n(t_n \mid t_1, \ldots, t_{n-1}, x_1 \ldots, x_{n-1}) \\
&\quad f_n(x_n \mid (t_1, x_1), \ldots (t_{n-1}, x_{n-1}), t_n) \\
&\quad \times S_{n+1}(T \mid t_1, \ldots, t_n, x_1 \ldots, x_n)
\end{aligned}$$

The main aim of the Proposition 4.3.14 was to make conditional the distribution of the current marks as time progresses, on marks and time points of all preceding events, i.e., the full set of the time points $(0, T)$, irrespective of the marks and their relative positions in time.

Another view to look at is through the hazard functions. Instead of specifying the conditional densities $p_n(\cdot \mid \cdot)$ as in (4.3.10) we express them in terms of their hazard functions

$$h_n(t \mid t_1, \ldots, t_{n-1}) = \frac{p_n(t \mid t_1, \ldots, t_{n-1})}{S_n(t \mid t_1, \ldots, t_{n-1})}$$

so that

$$p_n(t \mid t_1, \ldots, t_{n-1}) = h_n(t \mid t_1, \ldots, t_{n-1}) \exp\left(- \int_{t_{n-1}}^{t} h_n(u \mid t_1, \ldots, t_{n-1})\, du \right).$$

Using the conditioning in the hazard functions may now include the values of the preceeding marks as well as the length of the current and preceeding intervals. Thus, all the information is summarized in the internal history $\mathcal{H} \equiv \{\mathcal{H}_t : t \geq 0\}$ of the process and of this form the amalgam of hazard function functions and mark densities can be represented as a single composite function for the MPP as follows

$$\lambda^*(t, x) = \begin{cases} h_1(t) f_1(x \mid t) & (0 < t \leq t_1), \\ \vdots & \\ h_n(t \mid (t_1, x_1), \ldots, (t_{n-1}, x_{n-1})) \times & \\ f_n(x \mid (t_1, x_1), \ldots, (t_{n-1}, x_{n-1}), t) & (t_{n-1} < t \leq t_n,\ n \geq 2), \\ \vdots & \end{cases}$$

where $h_1(t)$ is the hazard function for the location of the initial point, $h_2(t \mid (t_1, x_1))$ the hazard function for the location of the second point conditioned by the location of the first point and the value of the first mark, and so on, while $f_1(x \mid t)$ is the density for the first mark given its location, and so on.

Hence a regular MPP N on $\mathbb{R}_+ \times \mathcal{X}$ can be represented piecewise by the function $\lambda^*(t, x)$ named *the conditional intensity function* with respect to its internal history $\mathcal{H}$.

Predictability, as it was defined in (4.3.5), is important in it that the hazard functions refer to the risk at the end of a time interval, not at the beginning of the next interval. Similarly, the conditional mark density refers to the distribution to be anticipated at the end of a time interval, not immediately after the next interval has begun.

Other form to write $\lambda^*(t,x)$ is through the $\mathcal{H}$-intensity of the ground process and the conditional density of a mark at t given $\mathcal{H}_{t-}$

$$\lambda^*(t,x)\,dtdx \approx \mathbb{E}\left[N(dt\times dx)\mid \mathcal{H}_{t-}\right] \approx \lambda_g^*(t)\,f^*(x\mid t)\,dtdx.$$

In addition to this, there are some remarks to do when the MPP has independent marks or unpredictable marks. These properties are resumed in the next proposition.

Proposition 4.3.15. *Define N to be a regular MPP on $\mathbb{R}_+\times\mathcal{X}$ with $\mathcal{H}$- intensity expressible as $\lambda^*(t,x)=\lambda_g^*(t)\,f^*(x\mid t)$, where $\lambda_g^*(t)$ is the $\mathcal{H}$-intensity of the ground process. Then N is*

1. *a process with independent marks if $\lambda_g^*(t)$ equals the $\mathcal{H}^g$-intensity for the ground process and $f^*(x\mid t)=f(x\mid t)$ for deterministic functions $\lambda(t)$ and $f(x\mid t)$.*
2. *a process with unpredictable marks if $f^*(x\mid t)=f(x\mid t)$ for deterministic functions $\lambda(t)$ and $f(x\mid t)$.*

Notice, that in a process with independent marks, the ground process and the marks are completely independent, whereas for a process with unpredictable marks, the marks can influence the subsequent evolution of the process, though the ground process does not influence the distribution of the marks.

Now, we show how these results can be used to estimate the likelihood function of a MPP in terms of its conditional intensity. Details and the proof can be found in Daley and Vere-Jones (2003).

Proposition 4.3.16. *Let N be a regular MPP on $[0,T]\times\mathcal{X}$ for some finite positive T and let $(t_1,x_1),\ldots,\left(t_{N_g(T)},x_{N_g(T)}\right)$ be a realization of N over the interval $[0,T]$. Then, the likelihood L of such a realization can be written as*

$$\begin{aligned} L &= \left[\prod_{i=1}^{N_g(T)}\lambda^*(t_i,x_i)\right]\exp\left(-\int_0^T\int_{\mathcal{X}}\lambda^*(u,x_i)\,du\ell_{\mathcal{X}}(d_{\mathcal{X}})\right) \\ &= \left[\prod_{i=1}^{N_g(T)}\lambda^*(t_i)\right]\left[\prod_{i=1}^{N_g(T)}f^*(x_i\mid t_i)\right]\exp\left(-\int_0^T\lambda_g^*(u)\,du\right), \end{aligned} \tag{4.3.11}$$

where $\ell_{\mathcal{X}}$ is the reference measure on $\mathcal{X}$.

Thus, the conditional intensity function with respect to the internal history $\mathcal{H}$ determines the probability structure of N uniquely.

This perspective of the likelihood is very natural and productive if one is concerned with developing new models. To elucidate the ideas proposed in this section we derived the maximum likelihood estimation for the generalized Hawkes-type cluster model and the ETAS model.

Example 4.3.17. Given a MPP with occurrence observations $(t_1, \ldots, t_n)$ for an interval $[0, T]$, marks $(x_1, \ldots, x_n)$, self-exciting function

$$\int_0^t \sum_k^K \phi_k (t-s)^{k-1} \exp\left(-\gamma\left(t-u\right)\right)\left(1+\delta x\right) ds$$

and conditional intensity

$$\lambda\left(t, x\right) = \frac{\tau + \phi w\left(t, x\right)}{\beta + \alpha w\left(t, x\right)} \left(1 + \xi \frac{x-u}{\beta + \eta w(t)}\right) \text{ for } x \geq u$$

the log-likelihood is given by equation (4.3.11)

$$l\left(;t_i, x_i\right) = \left[\sum_{i=1}^n \log \lambda^*\left(t_i\right)\right] + \left[\sum_{i=1}^n \log f^*\left(x_i \mid t_i\right)\right] - \left(\int_0^T \lambda_g^*\left(s\right) ds\right).$$

Now we calculate each sum separately in the discrete case. Denote $w\left(t, x\right) = \sum_{t<t_i} \sum_k^K \phi_k (t_i - t)^{k-1} \exp\left(-\gamma\left(t_i - t\right)\right)\left(1+\delta x_i\right)$. For the conditional intensity we have

$$\left[\sum_{i=1}^n \log \lambda^*\left(t_i\right)\right] = \sum_{i=1}^n \log\left\{\tau + w\left(t, x\right)\right\},$$

for the marks density

$$\left[\sum_{i=1}^n \log f^*\left(x_i \mid t_i\right)\right] = \sum_{i=1}^n \left(-\xi^{-1} - 1\right) \log\left(1 + \xi \frac{x_i - u}{\beta + \alpha w\left(t, x\right)}\right) - \log\left(\beta + \alpha w\left(t, x\right)\right)$$

and for the conditional intensity of the ground process

$$\left(\int_0^T \lambda_g^*\left(s\right) ds\right) = \tau n + \quad \left(1+\delta x_i\right) \sum_{k=0}^K \phi_k \left\{\int_0^T (T-s)^{k-1} \exp\left(-\gamma\left(T-s\right)\right) ds\right\}.$$

Note that $\int_0^T (T-s)^{k-1} \exp\left(-\gamma\left(T-s\right)\right) ds$ is the gamma function. Using integration by parts we get

$$\begin{aligned}\left(\int_0^T \lambda_g^*\left(s\right) ds\right) &= \tau n + \left(1+\delta x_i\right) \sum_{k=0}^K \phi_k \frac{\exp\left(-\gamma\left(T-t_i\right)\right)}{-\gamma} \\ &\quad \left\{\sum_{l=1}^k \frac{\Gamma\left(k+1\right)}{k\Gamma\left(k-l+1\right)\gamma^{l-1}} \left[1 - \left(T-t_i\right)^{k-l}\right]\right\}.\end{aligned}$$

In the same way we can derive the log-likelihood for the ETAS model.

Example 4.3.18. (Likelihood of the ETAS model)

Consider the ETAS Model proposed in Definition (4.3.9) with self-exciting function $w\left(t, x\right) = \int_0^t \left(1+\delta x\right)\left(1 + \frac{(t-s)}{\gamma}\right)^{-(1+\rho)} ds$ and conditional intensity as in the last example, the log-likelihood function for the MPP case is defined as

$$l\left(;t_i, x_i\right) = \left[\sum_{i=1}^n \log \lambda^*\left(t_i\right)\right] + \left[\sum_{i=1}^n \log f^*\left(x_i \mid t_i\right)\right] - \left(\int_0^T \lambda_g^*\left(s\right) ds\right).$$

Like the last example we calculate each term by separate

$$\begin{aligned} l\left(;t_i,x_i\right) &= \sum_{i=1}^{n}\log\left\{\tau+w\left(t,x\right)\right\}+\left[\sum_{i=1}^{n}\log f^{*}\left(x_i\mid t_i\right)\right] \\ &\quad -\left[\tau n+\phi\sum_{t_i<t}\left(1+\delta x_i\right)\frac{\gamma}{\rho}\left\{1-\left(1+\frac{\left(t-t_i\right)}{\gamma}\right)^{-\rho}\right\}\right]. \end{aligned}$$

Because the log-likelihood of these models is non-linear with respect to the parameters, the maximization of the log-likelihood is performed by using non-linear optimization techniques in R, R Development Core Team (2009).

4.4. The multivariate extension

Let now $\mathbf{X}_n=\{X_{1n},\ldots,X_{dn}\}$ be a stationary sequence multivariate random variables with distribution function $F\left(\mathbf{x}\right)$, $\mathbf{x}\in\mathbb{R}^d$. Assume that there exist sequences of normalizing d- variate sequences $\mathbf{a}_n$ and $\mathbf{b}_n\in\mathbb{R}^d$, and distribution function $G\left(\mathbf{x}\right)$ with nondegenerate marginals such that the limit (4.2.1) exists.

Furthermore, we assume that for sufficiently large thresholds u_j for $j=1,\ldots,d$, the marginal distribution of $X_j>u_j$ is GPD. Thus, it allows us to rewrite each marginal distribution j in the form of a self-excitement MPP as in the last section

$$F_j\left(x\right)=1-\left(\tau_j+\phi_j w_j(t)\right)\left(1+\xi_j\frac{x-u_j}{\beta_j+\eta_j w_j(t)}\right)_{+}^{-1/\xi_j}. \tag{4.4.1}$$

In the classical approach the results are based on the assumption that the multivariate distribution $F\left(\mathbf{x}\right)$ is in the domain of attraction of a MEVD $G\left(\mathbf{x}\right)$. The following proposition is essentially part *i)* of Proposition 5.15 in Resnick (1987).

PROPOSITION 4.4.1. *Let F_* be a multivariate distribution function with unit Frèchet marginals. Define $\mu\left(x_1,\ldots,x_d;\theta\right)=\mu\left(\left[\mathbf{0},\left(x_1,\ldots,x_d\right)\right]^c;\theta\right)$, then $F_*\in MDA\left(G_*\right)$ if and only if $1-F_*$ is regularly varying, i.e.,*

$$\lim_{t\to\infty}\frac{\log F_*\left(tx_1,\ldots,tx_d\right)}{\log F_*\left(t,\ldots,t\right)}=\frac{1-F_*\left(tx_1,\ldots,tx_d\right)}{1-F_*\left(t,\ldots,t\right)}=\frac{\log G_*\left(x_1,\ldots,x_d\right)}{\log G_*\left(1,\ldots,1\right)}=\frac{\mu\left(x_1,\ldots,x_d;\theta\right)}{\mu\left(1,\ldots,1;\theta\right)}, \tag{4.4.2}$$

where G_* is a MEVD with unit Frèchet marginals.

This Proposition suggests that for high thresholds the multivariate distribution function F can be approximated by its limit distribution function G. Two possible models for the joint tail F come from treating (4.4.2) as an identity for large t. We concentrate on the asymptotic dependent case, which gives a straightforward derivation. Substituting $s_j=tv_j$ in (4.4.2) and exploiting the homogeneity of $\mu\left(x1,\ldots,x_d;\theta\right)$ we have that for large s_j

$$1-F_*\left(s_1,\ldots,s_d\right)=\psi\left(t\right)\mu\left(s_1,\ldots,s_d;\theta\right), \tag{4.4.3}$$

where $\psi\left(t\right)=\frac{t\{1-F_*(t,\ldots,t)\}}{\mu(1,\ldots,1;\theta)}$. For convenience we choose F_* to have reciprocal uniform marginals, i.e., $F_{*,i}\left(x_i\right)=1-\frac{1}{x_i}$. In this case, since for $\mathbf{v}_i=\left(\infty,\ldots,\infty,v_i,\infty,\ldots,\infty\right)$ we have $F_*\left(\mathbf{v}_i\right)=F_{*,i}\left(v_i\right)$ and $\mu\left(\mathbf{v}_i;\theta\right)=\frac{1}{v_i}$ we get $\psi\left(t\right)=1$. Then, it follows that for s large we have

$$F_*\left(s_1,\ldots,s_d\right)\approx 1-\mu\left(s_1,\ldots,s_d;\theta\right).$$

Smith et al. (1997) used this approximation for multivariate thresholds exceedances. In our model the application is as follows.

Let $F(\mathbf{x})$ be a multivariate joint distribution, which marginals F_j for all $j=1,\ldots,d$ can be approximated as in (4.4.1). Define $S_j=\left(1-F_j\left(x_j\right)\right)^{-1}$. Thus, we get the next relation

$$F_*\left(x_1,\ldots,x_d\right)=F\left(S_1^{\leftarrow}\left(x_1\right),\ldots,S_d^{\leftarrow}\left(x_d\right)\right)$$

where $S_j^{\leftarrow}$ denote the inverse function of S_j.

Now suppose that $x_j>u_j$ for some marginal $j=1,\ldots,d$. replacing F_j by its marginal assumption in (4.4.1) we have;

$$S_j\left(x\right)=\left(\tau_j+\phi_j w_j(t)\right)^{-1}\left(1+\xi_j\frac{x-u_j}{\beta_j+\eta_j w_j(t)}\right)_+^{1/\xi_j}\quad\text{for } x_j>u_j. \tag{4.4.4}$$

Combining this with (4.4.3) we obtain the following approximation

$$F\left(x_1,\ldots x_d\right)=1-\mu\left(\left(\tau_1+\phi_1 w_1(t)\right)^{-1}\left(1+\xi_1\frac{x-u_1}{\beta_1+\eta_1 w_1(t)}\right)_+^{1/\xi_1},\ldots,\left(\tau_d+\phi_d w_d(t)\right)^{-1}\left(1+\xi_d\frac{x-u_1}{\beta_d+\eta_d w_d(t)}\right)_+^{1/\xi_d};\theta\right) \tag{4.4.5}$$

on $x_j>u_j$ for each j. Hence our approach is identical to that in the Smith et al. (1997) above the thresholds. Note that we take the thresholds to be fixed, which is consistent with existing practice.

4.4.1. Statistical inference. The multivariate extremes model is only valid when each variable Y_j exceeds some high threshold u_j for all $j=1,\ldots,d$. Therefore, the only relevant information for our model is if the observations occur above the threshold. Hence for the likelihood estimation, we considered the marginal observations as censored at the thresholds. Thus, the likelihood contribution for a typical observation $\left(x_1,\ldots,x_d\right)$ in which k components $j_1,\ldots,j_k$ exceed their thresholds is given by

$$f\left(x_1,\ldots,x_d\right)=\frac{\partial^k F(x_1,\ldots,x_d)}{\partial x_{j_1}\ldots\partial x_{j_k}}=\frac{\partial^k \mu(x_1,\ldots,x_d)}{\partial x_{j_1}\ldots\partial x_{j_k}}\times\prod_{l=j_1}^{j_k} s_l\left(x_l\right) \tag{4.4.6}$$

where $s_j\left(x_j\right)$ is the density function associated to the marginal distribution F_j for F defined as in (4.4.5). In our case it is

$$s_j\left(x_j\right)=\frac{\partial S_j\left(x_j\right)}{\partial x_j}=-\left(\tau_j+\phi_j w_j(t)\right)^{-\xi_j}\left(\beta_j+\eta_j w_j(t)\right)^{-1}S_j\left(x_j+u_j\right)^{1-\xi_j}$$

Example 4.4.2. We derive only the explicit form of the corresponding likelihood for the bivariate case without loss of the generality. The likelihood contribution corresponding to the point x_1,x_2 is denoted by $L_{\delta_1\delta_2}$, where δ_1 and δ_2 take the values 1 if the marginal contributes to the likelihood, i.e., $x_j>u_j$ or 0 if $x_j\leq u_j$ for $j=1,2$. In the d-variate case we have by inclusion-exclusion 2^d regions where we have to derive the likelihood contribution. In this example (4.4.6) gives

$$
\begin{aligned}
L_{00}(x_1,x_2) &= 1-\mu\left(\tau_1(t)^{-1},\tau_2(t)^{-1};\theta\right)\\
L_{10}(x_1,x_2) &= s_1(x_1)\frac{\partial\mu\left(S_1(x_1+u_1),\tau_2(t)^{-1};\theta\right)}{\partial x_1}\\
L_{01}(x_1,x_2) &= s_2(x_2)\frac{\partial\mu\left(\tau_1(t)^{-1},S_2(x_2+u_2);\theta\right)}{\partial x_1}\\
L_{11}(x_1,x_2) &= s_2(x_2)\,s_1(x_1)\frac{\partial^2\mu(S_1(x_1+u_1),S_2(x_2+u_2);\theta)}{\partial x_1\partial x_2}
\end{aligned}
$$

where $\tau_j(t)$ and $S_j(x_j)$ are defined as in (4.3.3) and (4.4.4).

The maximum likelihood estimation is straightforward in terms of its asymptotic properties and it is regular at interior points of the parameter space.

4.4.2. Models for the exponent measure μ. In this section we present some models to obtain a parametrization of the exponent measure μ together with some time-varying versions that have been considered in this work.

There are many families of extreme value distributions. We consider here only the bivariate case.

(1) *The asymmetric logistic* model (Galambos (1975),Tawn (1988a)) has the exponent measure

$$\mu(x_1,x_2;\psi_1,\psi_2,\alpha)=\frac{1-\psi_1}{x_1}+\frac{1-\psi_2}{x_2}+\left\{\left(\frac{\psi_1}{x_1}\right)^{1/\alpha}+\left(\frac{\psi_2}{x_2}\right)^{1/\alpha}\right\}^{\alpha},$$

where $\psi_1,\psi_2\in[0,1]$ and $\alpha\in(0,1]$. The density of the measure H can be derived as follows

$$h(v)=\{(\psi_2 v)^{\alpha}+(\psi_1(1-v))^{\alpha}\}^{1/\alpha-2}\{v(1-v)\}^{\alpha-2}(\psi_1\psi_2)^{\alpha}(1-\alpha).$$

We can show that the Logistic distribution has mass both in the interior and at the end points because $H(\{0\})=1-\psi_2$ and $H(\{1\})=1-\psi_1$. Independence is obtained when either $q\to 1$, $\psi_1=0$ or $\psi_2=0$ and total dependence correspond to $\psi_1=\psi_2=1$ and the limit $q\to 0$. A first special case for $\psi_1=\psi_2=1$ is the *symmetric logistic* model having all its mass in the interior. The second case is if $\psi_1=\psi_2=\alpha$ we have the *Gumbel* model

(2) The *Negative Logistic* model (*Joe (1990)*) takes the form

$$\mu(x_1,x_2;\psi_1,\psi_2,\alpha)=\frac{1}{x_1}+\frac{1}{x_2}+-\left\{\left(\frac{\psi_1}{x_1}\right)^{-1/\alpha}+\left(\frac{\psi_2}{x_2}\right)^{-1/\alpha}\right\}^{-\alpha}$$

where $\psi_1,\psi_2\in[0,1]$ and $\alpha>0$. This family is similar in structure to the logistic family with an symmetric version when $\psi_1=\psi_2=1$. The limiting cases are $\alpha\to\infty$ and $\alpha\to 0$ for the totally independent and totally dependent case, respectively.

(3) The *Polynomial* model (Klüppelberg and Mayer (2006)) is defined by means of Pickand's dependence functions $A(\cdot)$ as follows. Define an expansion polynomial

in $A(\cdot)$

$$A(w) = a_m w^m + a_{m-1} w^{m-1} + \cdots a_2 w^2 - \left(\sum_{k=2}^{m} a_k\right) w + 1$$

for $w \in [0,1]$ with $a_2 \geq 0$, $\sum_{k=2}^{m} a_k \geq 0$, $0 \leq \sum_{k=2}^{m} (k-1) a_k \leq 1$ and $\sum_{k=2}^{m} k(k-1) a_k \geq 0$. The corresponding exponent measure has $m-2$ parameters.

$$\mu(x_1, x_2; a_1, \ldots a_{m-2}) = \frac{1}{x_1} + \frac{1}{x_2} - \sum_{k=2}^{m} a_k \sum_{l=0}^{m-k} \begin{pmatrix} m-k \\ l \end{pmatrix} \frac{x_1^{l+k-1} x_2^{m-k-l-1}}{(x_1+x_2)^{m-1}}.$$

If we set $m = 5$, $a_5 = \psi_1/20$, $a_4 = \psi_2/12$, $a_3 = -(\psi_1 + \psi_2)/6$ and $a_2 = 1/2$, we have the *asymmetric mixed* modelled defined by Tawn (1988a) on terms of the exponent measure as follows

$$\mu(x_1, x_2; \psi, \alpha) = \frac{1}{x_1} + \frac{1}{x_2} + \left\{ \frac{(\alpha+\psi)x_1 + (\alpha + 2\psi)\, x_2)}{(x_1+x_2)^2} \right\},$$

where $\alpha \geq 0$, $\alpha + 2\psi \leq 1$,$\alpha + 3\psi \geq 0$. These constraints imply that $\psi \in [-0.5, 0.5]$ and $\alpha \in [0, 1.5]$. A *symmetric mixed* version is obtained when $\psi = 0$. In this model complete dependence cannot be obtained, independence is obtained when $\psi = \alpha = 0$.

The main assumption in these models is that the parameters associated to the structure of the dependence on each exponent measure, are invariant. However, in practice, one key issue is to state if we have a stationary exponent measure, on the whole sample. Few papers have tried to deal with that from of the point of view of exponent measures, but from the point of view of Copulas. See for example Patton (2006) who assumes that the functional form of the copula remains fixed over the sample whereas the parameters vary according to some evolution equation. Other alternative approachs may be considered to allow also for time variation in the functional form using a regime switching copula model, as in Rodriguez (2007).

Another point of view to look at is van den Goorbergh et al. (2005) proposes to model the time variation in the copula as a function of the conditional volatilities of the index returns. The motivation for this specification is the evidence on correlation breakdowns, which suggests that increased dependence occurs in hectic periods. An important drawback of these approaches is that they employ a fixed (parametric) structure for the pattern of changes in the copula parameter.

In order to reduce these limitations other authors as Dias and Embrechts (2004), Giacomini et al. (2007) have followed a semiparametric approach where the copula parameters are adaptively estimated in a time interval. The choice is performed via an adaptive estimation under the assumption of local homogeneity: for every time point there exists an interval of time homogeneity in which the copula parameter can be well approximated by a constant. This interval is recovered from the data using local change point analysis.

In this chapter we propose two types of models for the time-varying exponent measures functions. The first idea is to model by allowing the dependence parameter of an exponent measure function to envolve through time according to a combination of self-exciting functions. The second type of model capture local changes in the dependency structure of the exponent measure. The procedure adaptively estimates intervals of homogeneity, and for each interval an exponent measure function parameter is estimated.

4.4.2.1. *Parametric time-varying exponent measure functions.* In this models the dynamic in the structure of the exponent measure function $\mu(x_1, x_2; \theta)$ assumes that the dependence parameters θ evolves according to a particular self-exciting function, which like the models proposed for the marginals distributions can depend on the time and size of the exceedances over a threshold u.

There are many ways of capturing possible time variation in the conditional copula. The difficulty to specificy how the parameters evolve over time lies in defining a parametric evolution equation. We propose the following evolution equations for the bivariate symmetric logistic model.

Model 1. The first model is a natural extension inspired on a Taylor expansion of second order of an unknown function $f(y)$ which is related to the main dependence component of an exponent measure and y is defined as the absolute difference between the self-exciting functions of the marginal models $y = | w_1(t, x_1) - w_2(t, x_2) |$. Then, the evolution equation of the parameter $\alpha(t)$ takes form

$$\alpha(t) = \alpha_0 + \alpha_2 \nabla f + \alpha_2 \nabla^2 f.$$

Note the similarities of this model with the translogaritmic function, traditionally used in stochastic frontier analysis to approach production functions.

Model 2. The second model considers that there exists a direct relation between the correlation between the self-exciting functions of the marginal models

$$\alpha(t, x_1, x_2) = \alpha_0 + \alpha_1 \operatorname{corr}\left(w_1(t, x_1), w_1(t, x_2)\right).$$

To motivate this specification, we suppose that there is an increased dependence occurence in periods of extreme movements between the pairs. Hence, the self-exciting functions should exhibit some kind of correlation. If this is true, the theory predicts a negative value of α_1, by the fact that when the correlation augment $\alpha(t)$ should converge to zero.

4.4.2.2. *Time inhomogeneous exponent measure.* In this section we follow a semiparametric approach, since we will not specific a functional form for the parameter o the exponent measure. The idea is to select the time varying exponent measure parameters locally under the assumption of local homogeneity. This procedure was first proposed by Giacomini et al. (2007) for copulas models. The method bases on finding an interval of time homogeneity in which the exponent measure parameter can be well approximated by a constant.

In order to choose this interval of homogeneity we employ a local parametric test as introduced by Dias and Embrechts (2004).

Test of homogeneity. We suppose that we have d-variate multivariate extreme value distribution with n observations in the time interval $T = [0, n]$ and exponent measure $\mu(\mathbf{x}; \theta)$, where θ is a vector of parameters, that characterizes the dependence structure of the exponent measure μ. We are interested in one single change point analysis for θ in one time interval of size $m \subseteq T$. For instance, we test the null hypothesis $H_0 : \theta_{t0:m}$ versus the alternative $H_1 : \theta'_{t0:\delta^*} \neq \theta''_{\delta^*+1:m}$. Here $\delta*$ is the location time of the single change point if we reject the null hypothesis. The parameters $\theta_{t0:m}$, $\theta_{t0:\delta*}$ and $\theta'_{\delta*+1:m}$ are supposed to be unknown under both hypotheses.

The main idea is to test if these two sub samples $[t_0 : \delta*]$ and $[\delta * +1 : m]$ come from the same population under the assumption of that $\delta^* = \delta$ is known. The null hypothesis would be rejected for small values of generalized likelihood ratio test

$$\Lambda_\delta = \frac{\sup_\theta \prod_{t_0 \le i \le m} f\left(x_{i1}, \ldots, x_{id} \mid \theta\right)}{\sup_{\theta,\theta'} \prod_{t_0 \le i \le \delta} f\left(x_{i1}, \ldots, x_{id} \mid \theta'\right) \prod_{\delta < i \le m} f\left(x_{i1}, \ldots, x_{id} \mid \theta''\right)}, \tag{4.4.7}$$

where $f\left(x_{i1}, \ldots, x_{id} \mid \theta\right)$ is the density of $F\left(x_{i1}, \ldots, x_{id} \mid \theta\right)$.

Now, if we take logarithms we get

$$L_\delta\left(\theta'\right) = \sum_{t_0 \le i \le \delta} \log f\left(x_{i1}, \ldots, x_{id} \mid \theta'\right)$$

and

$$L_\delta\left(\theta''\right) = \sum_{\delta < i \le m} \log f\left(x_{i1}, \ldots, x_{id} \mid \theta''\right),$$

then the likelihood ratio equation (4.4.7) can be written as

$$-2\log\left(\Lambda_\delta\right) = 2\left\{L_\delta\left(\hat{\theta}'\right) + L_\delta\left(\hat{\theta}'\right) - L_m\left(\hat{\theta}\right)\right\}.$$

While δ is unknown , H_0 will be rejected for large values of

$$Z_m = \max_{t_0 \le \delta \le m} -2\log\left(\Lambda_\delta\right).$$

The asymptotic distribution of $\sqrt{Z_m}$ can be derived using extreme value theory, (see Embrechts et al. (1997)). In fact, if H_0 and all the necessary regularity conditions hold, we have

$$\lim_{m\to\infty} \mathbb{P}\left(A\left(\log\left(m\right)\right)\sqrt{Z_m} \le t - D_q\left(\log\left(m\right)\right)\right) = \exp\left(-2\exp\left(-t\right)\right) \tag{4.4.8}$$

for all t, where $A(x) = \sqrt{2\log(x)}$, $D_q(x) = 2\log(x) + \frac{q}{2}\log(\log(x)) - \log\left(\Gamma\left(\frac{q}{2}\right)\right)$, and $\Gamma(x) = \int_0^\infty s^{t-1}\exp(-s)\,ds$ is the gamma function.

Note, that the right hand side of (4.4.8) is the square of a Gumbel distribution function. It is known that the rate of convergence in results like (4.4.8) is very slow specially for small and moderate sample sizes. For this reason Gombay and Horváth (1996) proposed and Dias and Embrechts (2004) verified a result on which the test can yield better properties for smaller sample sizes.

Under H_0 and supposing that all necessary regularity conditions hold for $m \to \infty$

$$\mathbb{P}\left(\sqrt{Z_m} \ge x\right) \approx \frac{x^p}{2^{p/2}\Gamma\left(p/2\right)}\left\{\log u\left(h,l\right) - \frac{p}{x^2}\log u\left(h,l\right) + \frac{4}{x^2} + \mathcal{O}\left(x^{-4}\right)\right\}, \tag{4.4.9}$$

where $u(h,l) = (1-h)(1-l)/hl$, and h and l are functions depending on m for which Gombay and Horváth (1996) proposed the approximation $h(m) = l(m) = m^{-1}\log\sqrt[3]{m^2}$.

If we assume that there is exactly one change point in the exponent measure function, then the maximum likelihood estimation for the time of change is given by

$$\hat{\delta}_m = \min\left\{1 \le \delta < m : \; Z_m = -2\log\left(\Lambda_\delta\right)\right\}.$$

Define k as the minimal size of an interval for which all the necessary conditions of regularity and efficiency can be assumed.

In the case that there is no change, $\hat{\delta}_m$ will take a value near the limits of the sample. This holds because under the null hypothesis $\hat{\delta}_m/m \to \xi$ as $m \to \infty$ under regularity conditions, where $\mathbb{P}(\xi = 0) = \mathbb{P}(\xi = 1) = 1/2$ (see Csörgö and Horváth (1997)). The detection of several change-points in $d-$dimensional processes with unknown parameters can be done using so called binary segmentation procedure.

The method consists of the appliance of the likelihood ratio test for one change. If H_0 is rejected then we have the estimate of the time of the change $\hat{\delta_m}$. Next, we divide the sample into two subsamples $\left[1, \hat{\delta_m}\right]$ and $\left[\hat{\delta_m} + 1, n\right]$ and test H_0 for each one of them and so on. We continue this segmentation procedure until we do not reject the null hypothesis in any of the subsamples. However, under this procedure it is possible to provide the error probability of rejecting the hypothesis of homogeneity in more than one point.

Here, we follow an alternative method to avoid the estimation losses caused by the binary segmentation procedure. The method consists of the estimation of p minimal number of subsets for which the parameters of an exponent measure $\mu(x_1, x_2; \theta)$ can be approximated by a constant vector of parameters θ. The basic idea is to select a number of intervals $\mathcal{J}_j$ of length m_j for $j = 1, \ldots p$ in such a way that the time varying parameters $\theta(t)$ of an exponent measure $\mu(x_1, x_2 \mid \theta)$ can be approximated by constant vectors of parameters θ_j.

In order to find these intervals of local homogeneity we need to define first our null hypothesis to apply the test proposed in (4.4.9). The null hypothesis of homogeneity is that for which all the observations in the interval $\mathcal{J}_j$ follow the model with dependence parameter θ_j. The alternative hypothesis H_1 claims that there exist two types of internal subsets $\mathcal{I}_j\,[t_0, \delta_j]$ and $\mathcal{L}_j\,[\delta_j + 1, m_j]$ in $\mathcal{J}_j$ such that the parameters θ_j of the an exponent measure changes spontaneously in some point δ_j.

One possible setting to choose the parameters δ_j and m_j that have been successfully employed in this work is presented below.

Selection of the intervals $\mathcal{J}_j$. We define $\mathcal{I}_j := \mathcal{I}_{jq}\,[t, \delta_j]$ and $\mathcal{L}_j := \mathcal{L}_{jq\tau}\,[\delta_j + 1, m_j]$ with $\delta_j = t + qk$, $m_j = t + (q+1)k + \tau$, and $\mathcal{I}_{jq} \cup \mathcal{L}_{jq\tau} = \mathcal{J}_j$, where $\tau, q \in \mathbb{Z}^+$ are the largest integers for which the homogeneity in $\mathcal{J}_j$ is accepted, and k is the minimal number of observations for which all the necessary conditions of regularity and efficiency are assumed.

The formal procedure is as follows.

ALGORITHM 4.4.3. *Set $j = q = 1$ and $t = \tau = 0$.*

(1) *Test the null Hypothesis of homogeneity for the first interval $\mathcal{J}_1$, that is, all the observations in $\mathcal{J}_1 = \mathcal{I}_{11} \cup \mathcal{L}_{110}$ follow the model with dependence parameter θ_1. If this is rejected, then $\mathcal{J}_1 = \mathcal{I}_{11}$, and go to the step 3. Otherwise, increment q until homogeneity is rejected.*

(2) *If q is the integer for which homogeneity is rejected, set $\mathcal{I}_{0(q-1)}$ and increment τ until the largest possible in $\mathcal{L}_{1(q-1)\tau}$, for which $\mathcal{J}_1$ cannot be reject. The maximal estimated interval of homogeneity is $\mathcal{J}_1 = \mathcal{I}_{0(q-1)} \cup \mathcal{L}_{1(q-1)\tau}$.*

(3) *Finally, increment $j = j + 1$ and set $t_0 = k(q+1) + \tau$, $q = 1$ and $\tau = 0$. Repeat this procedure until $t_0 = n$.*

Note that the only critical parameter chosen is k. In our work k is the minimal number of exceedances, in at least a marginal of a $d-$dimensional extreme value model, such that the

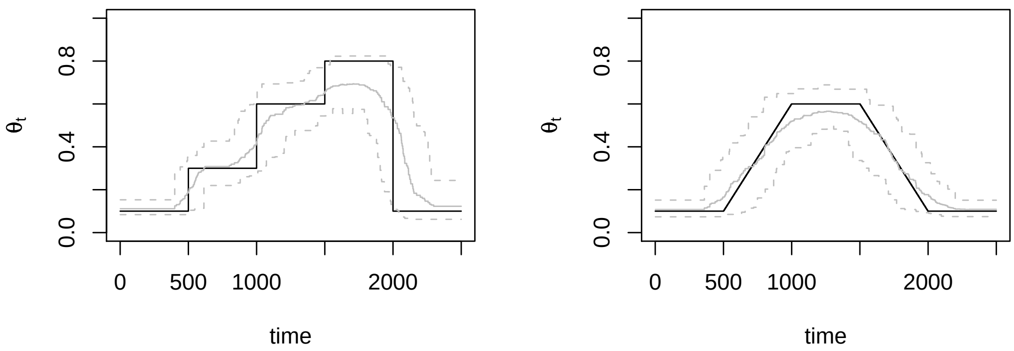

FIGURE 4.4.1. Time-inhomogeneous estimation of the pointwise median dependence parameter of the logistic exponent measure $\bar{\theta}$ (solid line) based on 100 simulations of the examples proposed, and their respective quantiles (dotted line) .

maximum likelihood can be consistently calculated. Small k leads to large standard deviations in the parameters θ, while an increase in k results in a more smoothed procedure.

Simulations. In this section we apply the test of homogeneity procedure on simulated bivariate data with dependence structure given by the logistic exponent measure $\mu\left(x_1, x_2 \mid \theta\right)$. In the two examples we generate 100 samples of length 2500. In the first example with 5 sudden jumps in the dependence parameter given by

$$\theta(t) := \begin{cases} 0.1 & 1 \leq t \leq 500 \\ 0.3 & 500 < t \leq 1000 \\ 0.6 & 1000 < t \leq 1500 \\ 0.8 & 1500 < t \leq 2000 \\ 0.1 & 2000 < t < 2500, \end{cases}$$

while in the second example the dependence parameter $\theta\left(t\right)$ has a smoother transition in the dependence parameter given by

$$\theta\left(t\right) := \begin{cases} 0.1 & 1 \leq t \leq 500 \\ 0.1 + \frac{(t-500)}{1000} & 500 < t \leq 1000 \\ 0.6 & 1000 < t \leq 1500 \\ 0.6 - \frac{(1500-t)}{2000} & 1500 < t \leq 2000 \\ 0.1 & 2000 < t < 2500. \end{cases}$$

Figure 4.4.1 shows the pointwise median and quantiles of the estimated parameter for $k = 100$ based on 100 simulations. Note that the estimations are more than satisfactory.

The difficulty of the procedure is principally due to the scars number of exceedances over the threshold u, in particular, when the data are asymptotically independent.

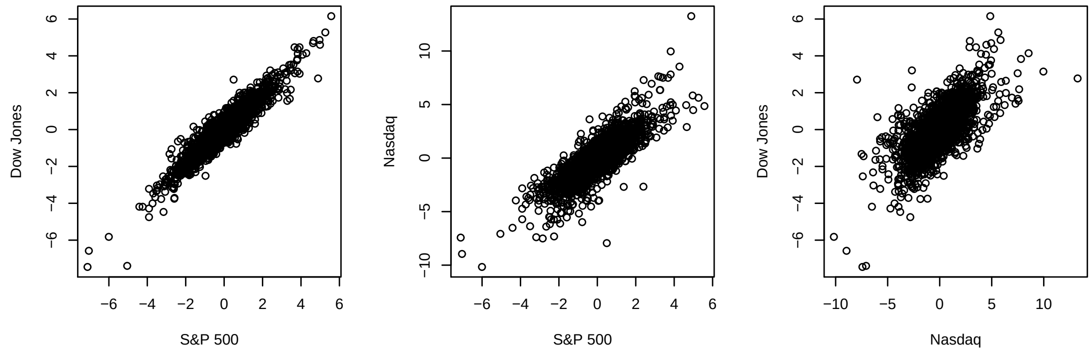

FIGURE 4.5.1. Bivariate scatter plots of the Dow Jones industrial average, NASDAQ and S&P500 returns for the period of investigation.

4.5. Empirical stress testing study in U.S. Future markets

There is documented evidence that the US market has the greatest influence on the other stock markets around the world (see for instance Poon et al. (2001)). Hence, an extreme event in the U.S. market is expected to have a major impact in the other markets around the world.

In this application we consider the three more importantly future contracts actually traded in U.S. markets: Dow Jones industrial average (DJI), NASDAQ index returns, and S&P500 index returns. A sample consisting of pairs of daily returns on the three indexes from January 1, 1990 to June 30, 2007 was obtained from Datastream.

This period of time saw the falling of world financial markets on the mini crash October 27,1998, following the bursting of the high-tech bubble, and the stock market fluctuations in September 11, 2001. Thus, this sample is perfectly suited to highlight the importance of the extreme dependence approach.

The bivariate log returns are plotted in Figure 4.5.1. From the plots, we see clearly that there are extreme observations in each sequence with a high dependence on the negative quadrant.

Table 4.A.1 in Appendix 4.A contains statistics of the data summarizing information about the unconditional distribution of the returns on the three indexes. The statistics show that all series exhibit skewness to the losses as well as excess kurtosis. For all returns we reject the null hypothesis of no serial correlation with p-values of zero for the three indexes. According to the results on the Engle's test, we also reject the null hypothesis of no heteroscedasticity for the three indexes. This anticipates the possible existence of volatility clusters in the three indexes.

4.5.1. Univariate self-exciting extreme value models. In order to investigate the multivariate extreme dependence among the returns we estimate first the self-exciting models for the univariate cases. The thresholds must be set high enough so that the maximum exceedances generalized Pareto. For this reason we choose to work with a 5% of the maxima of the sample.

The maximum likelihood estimates of the self-exciting extreme value models proposed in Section 4.3 for the marginal index return processes may be found in Tables 4.A.2, 4.A.3 and 4.A.4 in Appendix 4.A.

With the help of a likelihood Ratio test, we compare the different models and choose the best model for all indexes. The best fitted model is an ETAS model with predictable marks and influence of the same in the self-exciting function in three cases. In particular, the model gives evidence for the predictable marks (η) and influence of the marks in the self-exciting process. The assumption that the three index losses have heavy tails following a branching structure is strengthened by these results.

We provide also the W-statistics to assess our success in modelling the temporal behaviour of the exceedances of the threshold u. The W-statistics are defined to be

$$W = \xi^{-1} \ln \left(1 + \xi \frac{x-u}{\beta + \eta w\left(t\right)} \right).$$

This statistic states that if the GPD parameter model is correct, then the residuals are approximately independent unit exponential variables. The corresponding QQ-plots, displayed in Figure 4.5.2, do not show a substantial deviation from an exponential distribution.

To check that there is no further time series structure the autocorrelation function (ACF) for the residuals (left panel) and for their squares (right panel) are also included. Both autocorrelations are negligible at nearly all lags.

4.5.2. Empirical results for 1-dimensional value at risk (VaR) and expected shortfall (ES) estimations. We employ these estimated models to compute the risk measures VaR and ES. Not only *in sample* VaR and ES values are computed to examine the estimated model's ability but also *out of sample* the risk measures values are computed to evaluate the backtesting quality of the estimated model. The models are tested with a VaR and ES level α from 1% to 0.01%, i.e, a from an event in 100 days to an event in 40 years. As it is to be expected, if the measures were specified perfectly, the failure rate would equal to the prespecified level α. In other words, the more the failure rate approaches the pre specified level α, the more the risk measures VaR and ES helps investors to forecast their possible trading losses correctly.

The empirical results for the VaR and ES computations for S&P, NASDAQ , and Dow Jones are presented in Table 4.5.1 for some of the most important extreme events during the sample of study.

The first named the "mini-crash" was caused by an economic crisis in Asia. On Monday, October 27, the Dow Jones Industrial Average declined by 554.26 points (7.18%) to close at 7161.15. This represented the tenth largest percentage decline in the index since 1915. October 27 was also the first time that the cross-market trading halt circuit breaker procedures had been used since their adoption in 1988.

The second crash is attributed to the Russia defaults on the state short-term bonds, and devaluation of the rubble. The strong correction started in mid-August and it was not only specific to the U.S markets. Actually, it was much stronger in some other markets, such as the German market.

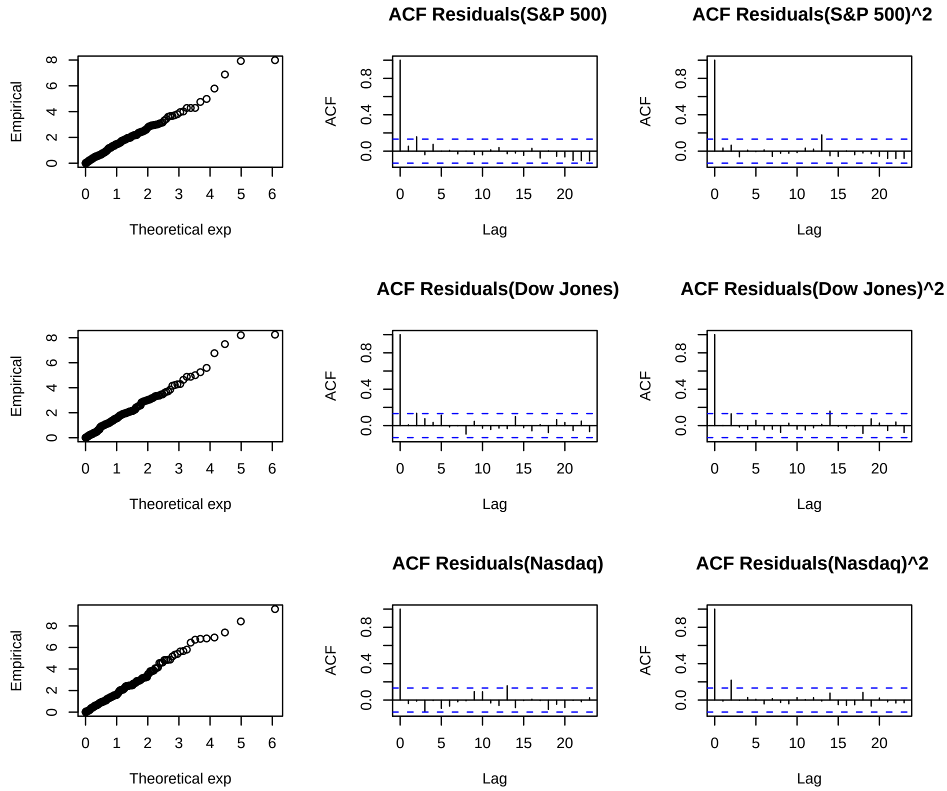

FIGURE 4.5.2. qq-plots of the residuals (left), autocorrelation function of the residuals (middle), autocorrelation function of the square of residuals (right), for the returns of the S&P 500 index (top panel), the Dow Jones index (middle panel) and the NASDAQ index (bottom panel).

The third to be mentioned is the NASDAQ crash in March and April, 2000. This index dropped precipitously between March 14, and April 14, about a cumulative loss of 50 % counted from its all-time high of 5,133 points reached on March 10, 2000. The drop was mostly driven by the so-called "New Economy" stocks which had risen nearly four-fold over 1998 and 1999 compared to a gain of only 50% for the S&P 500 index.

The fourth crash is associated to the attacks on the 11 September 2001. When the stock markets reopened on September 17, 2001, after the longest closure since the Great Depression in 1933, the three indexes together experienced more than 7 % day point decline. From the results of VaR and ES computations in Table 4.5.1 and Figure 4.5.3 we can come to the following conclusions. The dynamical estimation of the VaR measures for the three indexes varies considerably in relation to the number of crashes that already occurred. This happens because there are new information available to the self-exciting function which is responsible for the magnitude of the point process intensity. For instance, the probability of the extreme event in 1997 is lower than the probability of the Dot-com crash of 2001 since the information set, on which we condition, contains only past observations until 1990, so that the failure rates tend to be significantly different as the time go on.

Moreover, we test the null hypothesis of estimating correctly the Risk measures at time t_i against the alternative that the method systematically underestimates the returns $r_{t_{i+1}}$. Thus,

Time		S&P 500					
		$VaR_{0.01}$	$ES_{0.01}$	$VaR_{0.001}$	$ES_{0.001}$	$VaR_{0.0001}$	$ES_{0.0001}$
27.10.1997	7.113	2.507	3.207	4.140	5.055	6.276	7.474
31.08.1998	7.043	3.174	4.073	5.272	6.448	8.016	9.555
14.04.2000	6.004	3.4354	3.982	5.152	6.301	7.831	9.334
17.09.2001	5.046	3.8453	4.954	6.429	7.878	9.811	11.707
		Dow Jones					
		$VaR_{0.01}$	$ES_{0.01}$	$VaR_{0.001}$	$ES_{0.001}$	$VaR_{0.0001}$	$ES_{0.0001}$
27.10.1997	7.455	2.762	3.622	4.772	5.967	7.564	9.225
31.08.1998	6.578	3.549	4.705	6.256	7.868	10.022	12.262
14.04.2000	5.822	2.764	3.625	4.776	5.972	7.572	9.233
17.09.2001	7.396	3.959	5.276	7.037	8.866	11.312	13.853
		NASDAQ					
		$VaR_{0.01}$	$ES_{0.01}$	$VaR_{0.01}$	$ES_{0.01}$	$VaR_{0.001}$	$ES_{0.001}$
27.10.1997	7.426	2.107	2.691	3.462	4.141	5.038	5.826
31.08.1998	8.953	4.301	5.307	6.639	7.810	9.358	10.719
14.04.2000	10.168	8.429	10.517	13.280	15.708	18.919	21.7417
17.09.2001	7.077	6.115	7.600	9.565	11.292	13.576	15.584

TABLE 4.5.1. VaR and ES computations for S&P, NASDAQ , and Dow Jones for some of the most important extreme events during the sample of study.

the indicator for a violation at time t_i is Bernoulli $I_t := 1_{\{r_{t_{i+1}} > \{VaR_{\alpha,t_i}\}\}} \sim Be(1-\alpha)$. As it is described by McNeil and Frey (2000), I_t and I_s are independent for $t, s \in T$, then

$$\sum_{t_i \in T} \sim B(n, 1-\alpha). \tag{4.5.1}$$

Expression (4.5.1) is a two-tailed test that is asymptotically distributed as binomial. We perform the null hypothesis that it is a method that correctly estimates the risk measures against the alternative that the method has a systematic estimation error and gives too few or too many violations.

In all cases the models estimate correctly the conditional VaR for all the quantiles, the null hypothesis is rejected in Table 4.A.5 whenever the p-value of the binomial test is less than 1 percent.

Out-of-sample VaR computations is adopted to compare these VaR and ES with estimation sample we only know the "past" performance of these VaR models. However, to improve out estimation we need first to model the multivariate extreme dependence among the returns by means of exponent measures.

4.5.3. Multivariate estimation. Table 4.A.6 in Appendix 4.A reports parameter estimates for the bivariate asymmetric Logistic, the symmetric Logistic, the Negative Logistic and the polynomial model (with m = 5), when the dependency parameter is assumed to be constant over time. For all market pairs, the estimate of the dependency parameter is found to be positive and strongly significant. As expected, it is statistically and economically much larger for the pair S&P 500 and Dow Jones.

We also performed a likelihood ratio test to investigate whether a given exponent measure function is able to fit the dependence structure observed in the data. We found that the

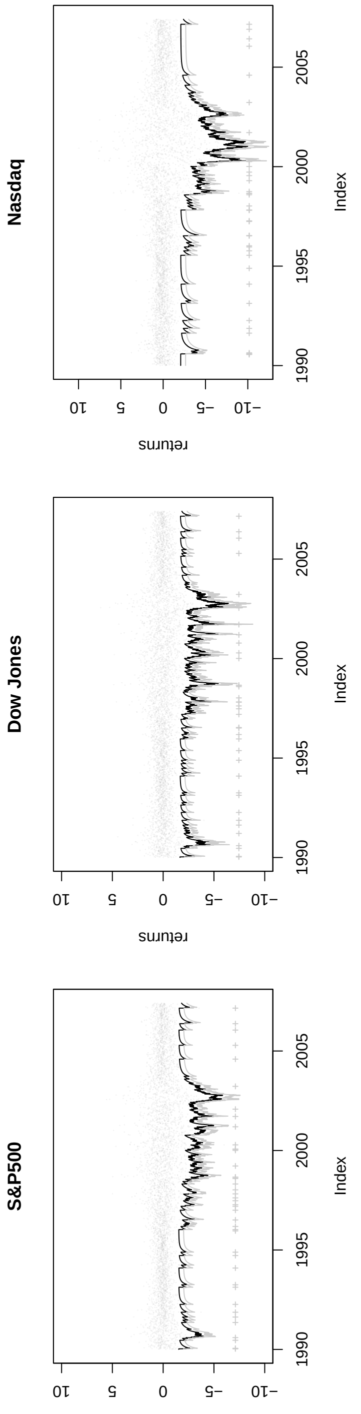

FIGURE 4.5.3. VaR (black line) and ES (gray line) computations at level $\alpha = 0.01$ (top panel) and $\alpha = 0.001$ (bottom panel) for S&P, NASDAQ and Dow Jones returns for some of the most important extreme events during the sample of study.

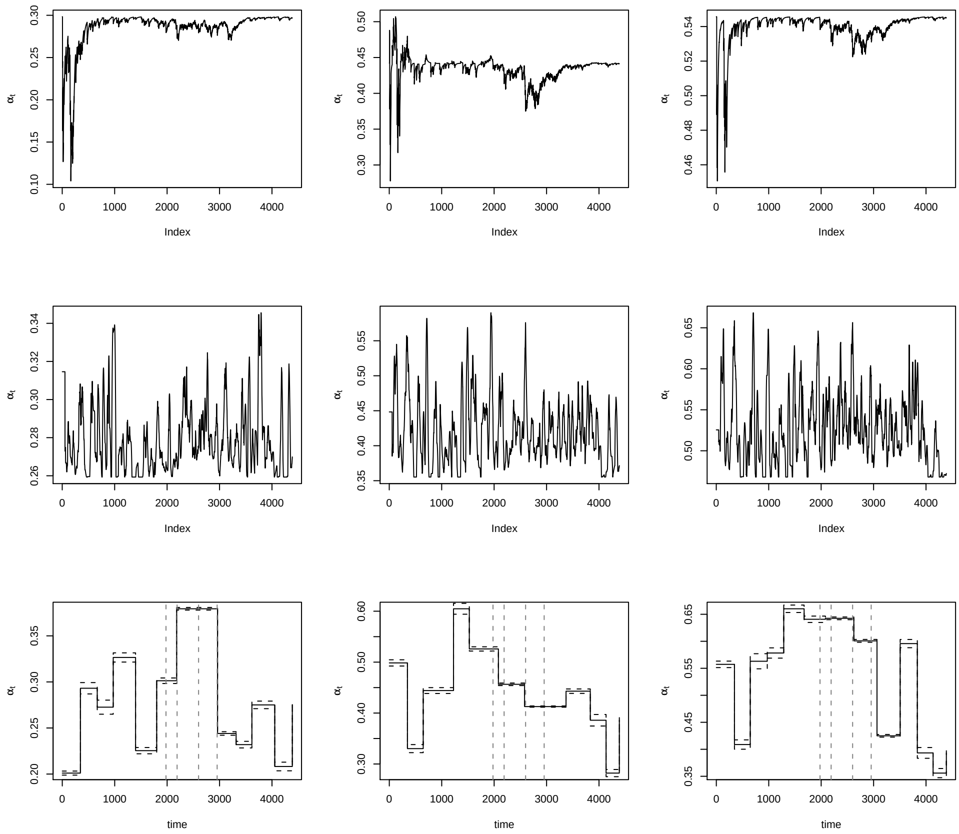

FIGURE 4.5.4. Estimated exponent measure dependence parameter $\alpha(t)$ for the pairs S&P 500 and Dow Jones (left), S&P 500 andNASDAQ (middle) , Dow Jones and NASDAQ (right) , by the parametric models 1 (top panel), 2 (middle panel), and time-inhomogeneous method (bottom panel) for $k = 150$.

symmetric Logistic model fits the data very well for the pairs S&P 500 and Dow Jones, S&P 500 and NASDAQ , and the asymmetric Logistic model for the pair Dow Jones and NASDAQ.

In the time-varying cases we turn our attention to the estimation of the parametric time-varying versions of only the best fitted bivariate static models. Due to the large number of possible models, we do not report the estimates for all combinations.

Table 4.A.7 in Appendix 4.A gives the results for the two parametric models proposed in Section 4.4.2.1 and the Figure 4.5.4 displays the evolution of the time varying coefficient α_t for the symmetric logistic case. A number of insights are possible from these results. Firsts, the model with a second order Taylor approximation gives the best maximum likelihood for the pairs S&P 500 and Dow Jones as well as S&P 500 and NASDAQ, while the pair Dow Jones and NASDAQ is better approximated by the second model with rolling Kendall tau correlation between the self-exciting functions of the univariate point processes. Thus, the time varying models yield a significantly better fit than the static case. However, not all of the parameters in the time varying second order Taylor approximation exponent measure were found to be significant and it is not clear that the functional form adopted should be the best approximation.

For these reasons, we use the adaptive estimation procedure based on the assumption of time inhomogeneous exponent measure estimated via Local Change Point procedure. Our goal in this study is to adopt different ways to deal with the dynamism of the extreme dependence. For the choice of the parameter k we take the minimal number of observations for which the maximum likelihood method can be used. In our application we adopted $k = 150$ for the three pairs under investigation. The results are depicted in Figure 4.5.4 and Table 4.A.7. In Figure 4.5.4 we displays in gray colour the time when most important events are taken place. Notice, that the local change point procedure predicts considerably well the times of these extreme events. For the first pair S&P 500 and Dow Jones the approach experiences three changes of the extreme dependence in relation to the crashes of the Asian crisis 1997, the Russian crisis 1998 and surprisingly the attacks of September 11, 2001. The first two of these crashes are at the same time the most important extreme events in the sample of study of our work.

In the case S&P 500 and NASDAQ the procedures spontaneously experienced two changes of dependence close to two of the most important extreme events between these two returns, the first happened after the Asian crisis in 1997 and the second in the Dot-com crash of 2000. In the last pair Dow Jones and NASDAQ only two of the crashes investigated here have a direct relation with a change of extreme dependence estimated with change point procedure. These extreme events are the Dot-com crash, 2000, and the attacks of the September 11, 2000. The estimation illustrates the improvement of the likelihood that can be obtained when one recognizes that there is a change-point in the dependence of the data and one is able to take this change into account when modelling the data. Following, we give three applications with the models estimated.

4.5.3.1. *Stress testing.* The extreme value theory is now familiar to practitioners. It allows, for example, to apply stress scenarios to a portfolio. However, the extension to the multivariate case is a difficult issue.

To illustrate how this result can be used for risk management, we consider an example which focuses on the extremes of the three bivariate pairs. Stress scenario in the bivariate case could be viewed as a failure set of the form

$$\mathcal{A} := \left\{(x_1, x_2) \in \mathbb{R}^2,\ \mathbb{P}\left(X_1 > x_1, X_2 > x_2\right) = \alpha\right\},$$

where α defines a contour failure area in the tail of the bivariate distribution. In our approach the bivariate probability can be written in terms of the exponent measure and the self-exciting point process approximation as follows

$$\begin{aligned}\mathbb{P}\left(X_1 > x_1, X_2 > x_2\right) &= 2 - \sum_{j=1}^{2} \left(\tau_j + \phi_j w_j(t)\right) \left(1 + \xi_j \frac{x_j - u_j}{\beta_{1j} + \eta_j w_j(t)}\right)^{-1/\xi_j} \\ &\quad -\mu\left(x_1, x_2; \theta\right), \text{ for } x_j \geq u_j.\end{aligned}$$

Note, that the probability α is associated with the waiting period for which the bivariate extremes can occur. To illustrate the concept of failure set with two variables, we provide different examples for the losses of returns of the three indexes for different probability scenarios and with different times where the extreme events happen.

For instance, we calculate the stress scenario for different levels of probability failure for the pair Dow Jones and NASDAQ for the Russian Crisis of 1998, the for the pair S&P 500 and NASDAQ in relation to the Dot-com crash of 2000, and for the pair S&P 500 and Dow Jones one week later to the attacks on September 11, 2001. The results are depicted in Figure 4.5.5. All the failure sets are calculated with the inhomogeneous exponent measure function calculated for each pair as in Figure 4.5.4. Notice, that the contours of probability failure are reduced when the pair of returns experience different extreme events. In fact, for the first case the Russian default of 1998, the probability of failure for this bivariate extreme event is circa of 37 years of waiting time.

A different result is given by the Dot-com crash of 2000. In this case this bivariate extreme event has a waiting time of approximately 13,9 years. This surprising and at the same time conservative result is first due to strong movements that others experienced during the week before by the NASDAQ index made a cumulative losses of 21,8 and second, to the branching structure of the self-exciting point process approach to deal with the cluster of the extremes.

Finally, for the extreme events related to the attacks on September 11, 2001, we found that this events should happen each 5,5 years. Another point of view to look at these extreme events was done by Straetmans et al. (2008) who derived non-parametric estimates for co-crash and co-boom probabilities of sectoral indexes conditional on simultaneous meltdowns in a market portfolio index. The main difference with our approach is the dynamical character of the extreme dependence.

4.5.3.2. *Stress testing in higher dimensions.* Stress testing becomes intractable in higher dimensions because the number of points of the failure sets increases very quickly. Moreover, the probability failures are not more contour lines instead of this, they are surfaces in the trivariate case for example. Thus, an alternative is to simulate different short positions of portfolios with different weights.

To illustrate the Monte-Carlo applications, we will trivariate failure sets under the hypothesis of a trivariate logistic dependence function (see Example 2.3.10). The failure set for this example is defined as

$$\begin{aligned}\mathcal{A}: \; &= \; \big\{(x_1, x_2, x_3) \in \mathbb{R}^3, \; \mathbb{P}\left(X_1 > x_1, X_2 > x_2, X_3 > x_3\right) \\ &\quad : X_1 w_1 + X_2 w_2 + X_3 w_3 \geq q\big\} \text{ for all } x_j \geq u_j,\end{aligned}$$

where w_1, w_2 and w_3 are weights of the returns associated to each market subject to that the sum of weights is equal 1 and q is the accumulative loss for a choosing position.

We can compute the probability of having large losses lower than 12 on a Monte Carlo simulation of 30.000 weighed portfolios for a short position in the Dot-com crash on April 14, 2000. The results of the simulation are displayed in Figure 4.5.6. We differentiated between portfolios where a weight w_i is larger than 0.5 for a marginal in comparison to the others returns.

Some interesting results can be concluded from these figures. First, portfolios where all the weights are lower than 0.5 show a good degree of diversification and they are relatively homogeneous for different combinations of weights. Furthermore, there exists a proportional relation between the probability of failure α and the losses of the weighed portfolio.

Second, weights larger then 0.5 for the S&P 500 ($w_1 > 0.5$) and Dow Jones index ($w_2 > 0.5$) show the higher degree of diversification. However, on contrary to the Dow Jones, the S&P 500 index has not lower losses positions with a lower probability of failure.

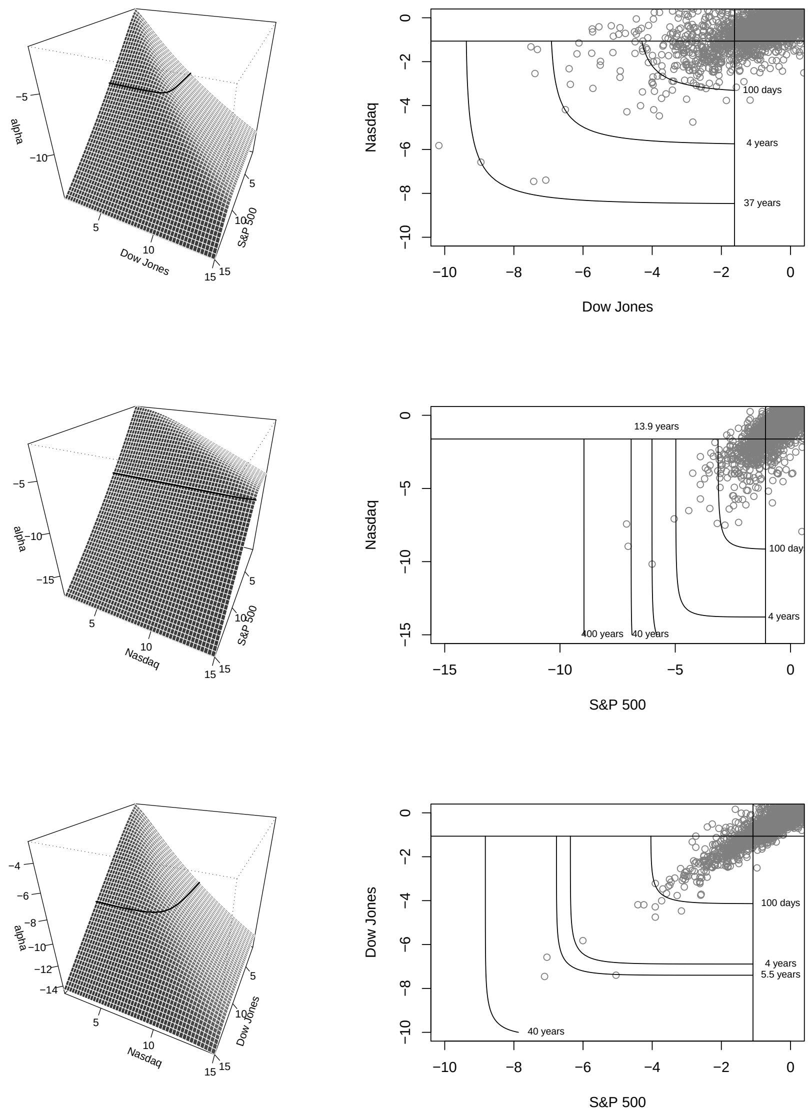

FIGURE 4.5.5. Perspective densities in logaritmic scale (left) and contour plots of failure sets (right) for the pairs S&P 500 and Dow Jones (top panel), S&P 500 and NASDAQ (middle panel), Dow Jones and NASDAQ (bottom panel) for different waiting times or failure probabilities.

Finally, the weighted portfolios where the weight of the NASDAQ return is larger than 0.5 ($w_3 > 0.5$) show heterogeneous results in the short positions of the losses. In addition to this, the most dangerous portfolios or the worth cases of the Monte Carlo simulation is where the weigh of the NASDAQ return is larger than 0.5. For instance, it is possible to find portfolios with losses of the order of 10 with a probability of 0.01 or equivalently an event in 100 days!

4.5.3.3. *The worst case scenario for the VaR.* Due to its simplicity, but also because of regulatory reasons, Value at Risk remains one of the most popular risk measures. The aim of this section is to give more insight into the problem of managing the VaR of a joint position resulting from the combination of different dependent risks. In particular the problem

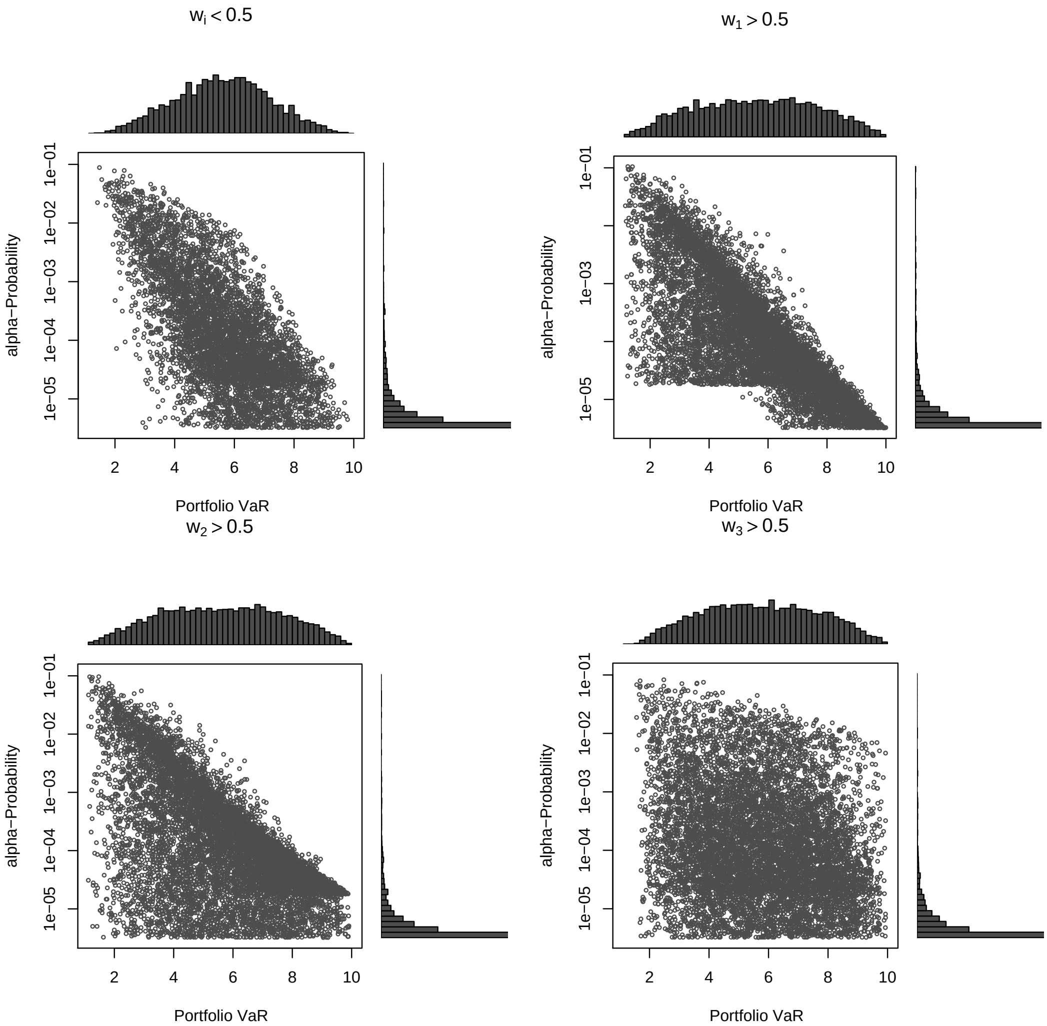

FIGURE 4.5.6. Monte Carlo simulation of 30.000 weighed portfolios for a short position in the Dot-com crash on April 14, 2000. (Top-left) Portfolios where all the weights are lower than 0.5 ($w_i < 0.5$). (Top-right) Portfolios where the weigth of the S&P 500 index is larger then 0.5 ($w_1 > 0.5$). (Bottom-left) Portfolios where the weigth of the Dow Jones index is larger then 0.5 ($w_2 > 0.5$). (Bottom-rigth) Portfolios where the weigth of the NASDAQ index is larger then 0.5 ($w_3 > 0.5$).

of finding the best possible lower bound or equivalent to find the worst possible Value-at-Risk (VaR) for a corresponding aggregate position of a portfolio. This problem has received a considerable interest in Finance, Risk management and Insurance mathematics; see an introduction in Embrechts et al. (2003). Only partial results have been obtained for the 2-dimensional case, when no information on the structure of dependence of the random vector is available.

In this section, our goal is to address some of the issues outlined above in a numerically tractable way. To this end we introduce the notion of worst-case VaR as the following failure set in the 3-dimensional case:

$$\begin{aligned} \mathcal{A}: &= \left\{(x_1, x_2, x_3) \in \mathbb{R}^3,\ \mathbb{P}\left(X_1 > x_1, X_2 > x_2, X_3 > x_3\right) = \alpha \right. \\ &: \quad \left. \alpha = \arg\max\, VaR_\alpha\left(X_1 + X_2 + X_3\right)\right\}, \end{aligned}$$

This set can be interpreted as given a failure probability α, which is the worst case such that the cumulative losses from $X_1 + X_2 + X_3$ is maximized. Another way to look at is, it is to try to minimize the waiting time (maximize the failure probability α) for which these cumulative losses could happen.

In this approach we assume that the true distribution of returns is partially known and can be modelled through the self-exciting models and the dependence among the marginals is well modelled by means of a Logistic exponent measure.

We apply this scenario to each one of the most important extreme events in the sample period. We simulate 30.000 portfolios for different magnitudes of losses in each marginal. The results are depicted in Figure 4.5.7. At difference to the stress scenario presented in the last section, the weighs w_i are the proportion of the losses in each marginal in relation to the sum of all losses.

For the first extreme event, the mini crash produced by the Asian crisis, the VaR for portfolios without proportions larger than 0.5, i.e., $w_i < 0.5$ for all the marginals, the Worst case for the 30.000 simulations is comparably higher than for other classes of portfolios where the proportions is larger than 0.5 at least in a marginal. In fact, the lowest failure probabilities are for portfolios where a component, the S&P 500 or the Dow Jones, is larger than 0.5.

The worst case for the Russian default corresponds to portfolios where the proportions are lower than 0.5 for all marginals. In second place for portfolios where the main marginal is the NASDAQ return. In third place for portfolios where the main component is the Dow Jones index. The best cases among these worst cases are for portfolios composed mostly by the S&P 500 index.

A different result was found for the Dot-com market crash in year 2000. The Worst cases are clearly those where NASDAQ return is the most import part of a portfolio, with some portfolios where the waiting time (or failure probability) does not exceed the 100 days ($\alpha = 0.01$). For portfolios where the major component of losses is given by the S&P 500 or the Dow Jones returns the failure probability is indistinguishable until losses of order 10. However, for losses bigger then 10, losses where the major component are the Dow Jones returns, that have a higher failure probability than losses where the major component are S&P 500 the returns.

Finally, during the attacks of September 11, 2001, the simulations of the worst case are not too different from the simulations of the Dot-com crash. The most important difference is that in this case the worst cases are those where the NASDAQ is the main component or where the proportions for each marginal are lower than 0.5. There is a small difference for portfolios where the NASDAQ is the main component from portfolios where the losses are bigger than 10. Continuing with the portfolios where S&P 500 or Dow Jones returns are the main component, the worst cases are also distinguishable for losses bigger than 10 where the Dow Jones returns have a higher failure probability.

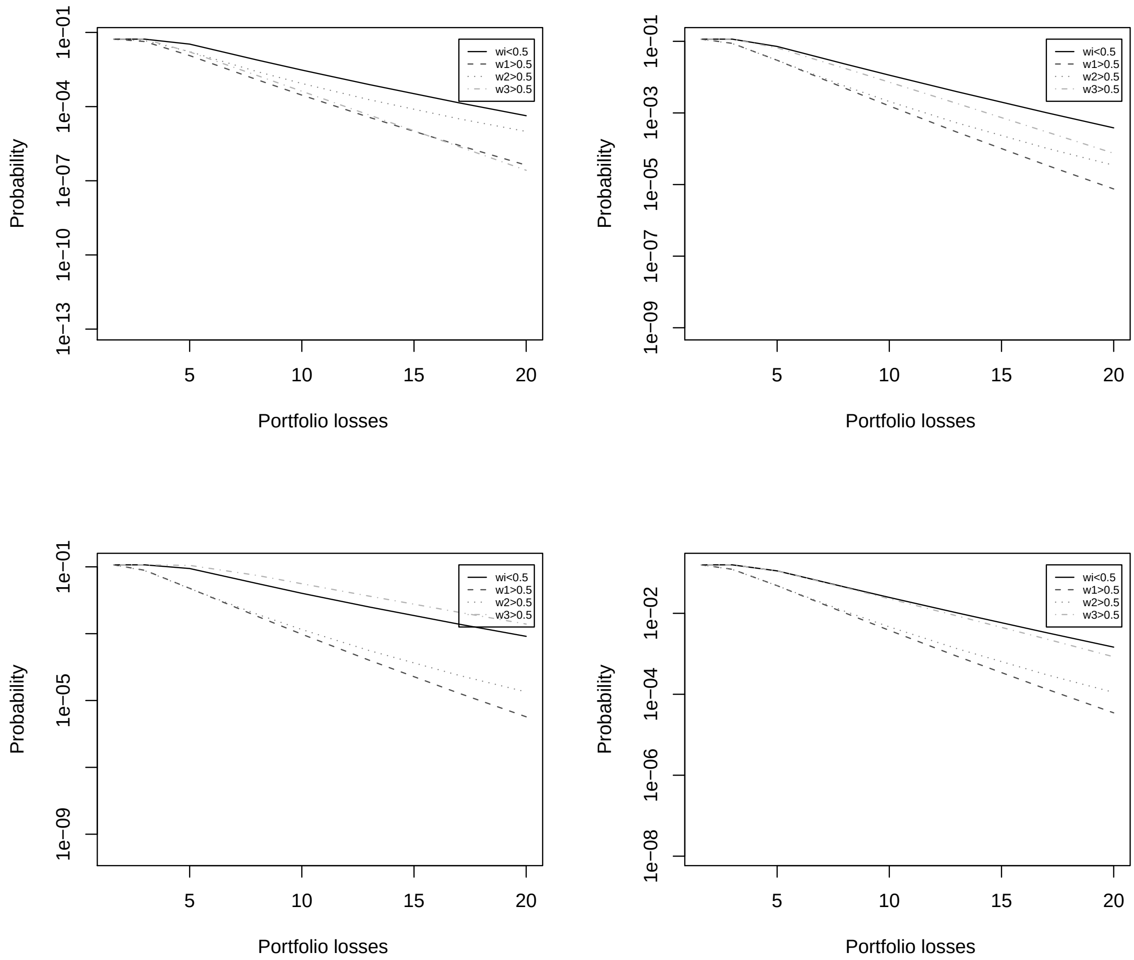

FIGURE 4.5.7. The Worst case scenario simulation for the VaR for the most important extreme events during the sample period. (Top-left) The Asian crisis. (Top-right) The Russian default. (Bottom-left) The Dot-com market crash. (Bottom-right) the Attacks of September 11, 2001.

4.6. Conclusions

In this chapter we introduced models of self-exciting point process to model the extreme events of stationary sequences no iid as it is the case of the most financial returns as a marked point process. We observe that under this methodology the estimation of such models can be forwardly derived through conditional intensities having the Jannosy measures as base of them.

Different models were proposed having in mind the simplicity of the structure of the self-exciting functions. However, other more complicate structures could be adopted. Nonetheless, the models used and their estimations in different U.S stock markets were more than satisfactory. In average, all the models fit the VaR and ES well, i.e., in terms of capital requirement; the models keep necessary capital to guarantee the desired confidence level.

Furthermore, in the multivariate case we discussed extensions of existing results on multivariate extreme dependence for allowing time-varying models, and employed it to construct flexible models. The results suggest that changes in the time dependence should be considered to make an accurate estimation, where the semiparametric model gives the best fit.

We show how this methodology may also be used in several contexts, such as conditional asset allocation or Value-at-Risk computation in a non-Gaussian framework. We apply different scenarios to each one of the most important extreme events in the sample period. We simulate portfolios for different magnitudes of losses in each marginal. The results are mixed and they depend on the time and on the pair for which they were calculated. Another issue to mention it is the dependence structure between the data is considered to be symmetric and it is modelled with only one parameter. This assumption is very strong, as the dependence between different pairs of data might be different. A much broader class of multivariate extreme models should be tested.

4.A. Tables and Figures

Indexes	mean	sd	min	max	skewness
S&P 500	0.033	0.988	-7.112	5.574	-0.1154
Dow Jones	0.0359	0.975	-7.454	6.154	-0.239
NASDAQ	0.0394	1.491	-10.168	13.254	-0.022
Indexes	**kurtosis**	**Box.test**	**Jarque-Bera**	**ADW**	**Engle (10)**
S&P 500	3.968	12.264**	2894.325 ***	-16.134***	502.746***
Dow Jones	4.776	12.099**	4219.09***	-16.452***	465.507***
NASDAQ	6.027	16.430*	6652.893***	-14.605***	851.999***

TABLE 4.A.1. Summary statistics for the stock index returns. Asymptotic p-value are shown in the brackets. *,**,*** denote statistical significance at the 1, 5 and 10 % level respectively. The Ljung-Box test statistic for serial correlation up to the 5-th order.

Models	τ	ϕ	γ	ρ	δ	ξ	β	η	*l.like*
Model a	0.0085	0.0247	101.4323	3.1633	0.0727	0.0198	0.0617		-921,976
	(0.0026)	(0.0082)	(119.8554)	(3.1821)	(0.2683)	(0.0702)	(0.0599)		
Model b	0.0089	0.0283	72.7049	2.6500	0.0882	0.1168	0.4191	0.0598	-917,324
	(0.0027)	(0.0089)	(74.9103)	(2.2922)	(0.2547)	(0.0673)	(0.0730)	(0.0286)	
Model c	0.0090	0.0216	0.0266		0.0166	0.1098	0.6169		-922,565
	(0.0026)	(0.0066)	(0.0068)		(0.2410)	(0.0702)	(0.0599)		
Model d	0.0096	0.0247	0.0311		0.0270	0.1180	0.4234	0.0513	-918,1136
	(0.0026)	(0.0069)	(0.0075)		(0.2238)	(0.0674)	(0.0730)	(0.0234)	
Model e	0.0096	0.0246	0.0311		0.0338	0.1180	0.4235	0.0510	-918,111
	(0.0026)	(0.0071)	(0.0075)		(0.2282)	(0.0674)	(0.0730)	(0.0235)	
Model f	0.0096	0.7947	0.0311		0.0270	0.1180	0.4234	1.6504	-918,1136
	(0.0026)	(0.1445)	(0.0075)		(0.2239)	(0.0674)	(0.0730)	(0.6421)	

TABLE 4.A.2. Results for the S&P 500 returns. *Model a* is an ETAS model without predictable marks ($\eta = 0$) (Equation 4.3.9). *Model b* is an ETAS model (Equation 4.3.9). *Model c* is a simple Hawkes model without predictable marks (Equation 4.3.7). *Model d* is a simple Hawkes model (Equation 4.3.7). *Model e* is a generalized Hawkes model (Equation 4.3.6). *Model f* is an exponential Hawkes model (Equation 4.3.8). Standard deviations are shown in between parentheses. *l.likes* are the results of the maximization of the log-likelihood estimation

Models	τ	ϕ	γ	ρ	δ	ξ	β	η	*l.like*
Model a	0.0125	0.0143	45.7714	1.4238	0.9112	0.1777	0.5910		-962.197
	(0.0038)	(0.0076)	(43.3483)	(1.0451)	(0.8951)	(0.0784)	(0.0607)		
Model b	0.0135	0.0232	31.5628	1.3525	0.4963	0.1428	0.3762	0.0811	-955.417
	(0.0039)	(0.0096)	(26.0828)	(0.8951)	(0.4936)	(0.0758)	(0.0659)	(0.0451)	
Model c	0.0133	0.0117	0.0218		0.5115	0.1777	0.5910		-963.750
	(0.0037)	(0.0059)	(0.0075)		(0.5815)	(0.0784)	(0.0607)		
Model d	0.0155	0.0196	0.0341		0.2855	0.1488	0.3749	0.0712	-957.224
	(0.0038)	(0.0069)	(0.0104)		(0.3504)	(0.0755)	(0.0654)	(0.0353)	
Model e	0.0153	0.0211	0.0341		0.1489	0.1475	0.3732	0.0776	-957.348
	(0.0038)	(0.0062)	(0.0102)		(0.1431)	(0.0753)	(0.0652)	(0.0340)	
Model f	0.0155	0.5733	0.0341		0.2855	0.1488	0.3749	2.0864	-957.224
	(0.0038)	(0.1506)	(0.0104)		(0.3504)	(0.0755)	(0.0654)	(0.8349)	

TABLE 4.A.3. Results for the Dow Jones returns. *Model a* is an ETAS model without predictable marks ($\eta = 0$) (Equation 4.3.9). *Model b* is an ETAS model (Equation 4.3.9). *Model c* is a simple Hawkes model without predictable marks (Equation 4.3.7). *Model d* is a simple Hawkes model (Equation 4.3.7). *Model e* is a generalized Hawkes model (Equation 4.3.6). *Model f* is an exponential Hawkes model (Equation 4.3.8). Standard deviations are shown in between parentheses. *l.likes* are the results of the maximization of the log-likelihood estimation

Models	τ	ϕ	γ	ρ	δ	ξ	β	η	*l.like*
Model a	0.0044	0.0165	202.7318	4.5535	0.2132	0.1123	1.0037		-959.6560
	(0.0017)	(0.0072)	(76.7710)	(5.2351)	(0.2512)	(0.0792)	(0.1041)		
Model b	0.0050	0.0211	87.4314	2.8092	0.3316	0.0654	0.5673	0.0703	-948.9087
	(0.0017)	(0.0079)	(82.2368)	(2.1316)	(0.2685)	(0.0690)	(0.1035)	(0.0321)	
Model c	0.0046	0.0148	0.0195		0.1762	0.1123	1.0037		-960.0277
	(0.0017)	(0.0063)	(0.0054)		(0.2335)	(0.0792)	(0.1041)		
Model d	0.0053	0.0192	0.0272		0.2362	0.0700	0.5663	0.0643	-949.7563
	(0.0017)	(0.0067)	(0.0065)		(0.2142)	(0.0694)	(0.1040)	(0.0280)	
Model e	0.0053	0.0207	0.0274		0.1367	0.0701	0.5641	0.0698	-949.8990
	(0.0017)	(0.0062)	(0.0065)		(0.0900)	(0.0695)	(0.1041)	(0.0271)	
Model f	0.0053	0.7062	0.0272		0.2362	0.0700	0.5663	2.3611	-949.7563
	(0.0017)	(0.1466)	(0.0065)		(0.2142)	(0.0694)	(0.1040)	(0.7307)	

TABLE 4.A.4. Results for the NASDAQ returns. *Model a* is an ETAS model without predictable marks ($\eta = 0$) (Equation 4.3.9). *Model b* is an ETAS model (Equation 4.3.9). *Model c* is a simple Hawkes model without predictable marks (Equation 4.3.7). *Model d* is a simple Hawkes model (Equation 4.3.7). *Model e* is a generalized Hawkes model (Equation 4.3.6). *Model f* is an exponential Hawkes model (Equation 4.3.8). Standard deviations are shown in between parentheses. *l.likes* are the results of the maximization of the log-likelihood estimation

Time	$VaR_{0.01}$	$VaR_{0.001}$	$VaR_{0.0001}$
S&P 500	42 (0.426)	7 (0.922)	1 (0.928)
Dow Jones	38 (0.209)	7 (0.922)	0 (0.645)
NASDAQ	38 (0.209)	7 (0.922)	1 (0.928)

TABLE 4.A.5. Test hypothesis of estimating correctly the Risk measures at time $t_i$against the alternative that the method systematically underestimates the returns $r_{t_{i+1}}$. The indicator for a violation at time t_i is Bernoulli $I_t := 1_{\{r_{t_{i+1}} > \{VaR_{\alpha,t_i}\}\}} \sim Be(1-\alpha)$. As it is described by McNeil and Frey (2000), I_t and I_s are independent for $t, s \in T$, then $\sum_{t_i \in T} \sim B(n, 1-\alpha)$ is a two-tailed test that is asymptotically distributed as binomial.

	α	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	max.ll
			Symmetric Logistic					
S&P - Dow Jones	0.279 (0.009)							-2155.519
S&P - NASDAQ	0.423 (0.012)							-2709.278
Dow Jones - NASDAQ	0.536 (0.014)							-2913.327
			Asymmetric Logistic					
S&P - Dow Jones	0.254 (0.008)	0.982 (0.010)	0.999 (0.001)					-2156.921
S&P - NASDAQ	0.419 (0.013)	0.992 (0.010)	0.999 (0.002)					-2708.89
Dow Jones - NASDAQ	0.509 (0.017)	0.819 (0.035)	0.999 (0.001)					-2918.617
			Polynomial					
S&P - Dow Jones	0.999 (0.001)							-2425.674
S&P - NASDAQ	0.994 (0.001)							-2778.583
Dow Jones - NASDAQ	0.940 (0.041)							-2919.271
			Negative Logistic					
S&P - Dow Jones	0.961 (0.075)	0.787 (0.035)	0.999 (0.001)					-2736.025
S&P - NASDAQ	0.681 (0.026)	0.999 (0.001)	0.893 (0.037)					-2994.183
Dow Jones - NASDAQ	1.766 (0.109)	0.270 (0.017)	0.992 (0.02)					-3392.393

TABLE 4.A.6. Parameters estimates for the bivariate asymmetric Logistic, the symmetric Logistic, the Negative Logistic and the Polynomial model, when the dependency parameter is assumed to be constant over time.

	α	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6
	Parametric Time - varying exponent measures (Model 1)						
S&P - Dow Jones	0.298 (0.011)	1.263 (1.308)	-0.439 (2.018)	0.013 (7.907)	0.016 (1.824)	0.028 (0.733)	-2148.438
S&P - NASDAQ	0.442 (0.017)	-0.964 (0.677)	3.211(2.003)	0.008 (0.001)	0.002 (0.001)	0.007 (1.307)	-2707.949
Dow Jones - NASDAQ	0.545 (0.020)	0.026 (0.757)	0.870 (2.006)	0.015 (2.145)	0.002 (0.003)	0.010 (4.412)	-2913.074
	Parametric Time - varying exponent measures (Model 2)						
S&P - Dow Jones (w = 52)	0.336 (0.039)	-0.076 (0.051)					-2154.354
S&P - NASDAQ (w = 55)	0.516 (0.029)	-0.161 (0.043)					-2701.994
Dow Jones - NASDAQ (w = 46)	0.592 (0.023)	-0.124 (0.040)					-2908.562
	Parametric Time - varying exponent measures (Model 2)						
S&P - Dow Jones (k = 150)							-2146.431
S&P - NASDAQ (k = 150)							-2694.732
Dow Jones - NASDAQ (k = 150)							-2903.452

TABLE 4.A.7. Parameters estimates for the bivariate asymmetric Logistic, the symmetric Logistic, the Negative Logistic and the Polynomial model, when the time-variation adopts a parametric or semiparametric form.

CHAPTER 5

On the estimation of M_4 processes through mixture Dirichlet process models

"The infinite in mathematics is always unruly unless it is properly treated."
(Edward Kasner and James Newman)

5.1. Introduction

The difficulty for statistical applications of mutivariate extreme value modes is when the number of dimensions d increase considerably, these do not reduce to a finite dimensional parametric family. Accordingly, there is potential explosion in the class of models to be considered. Nowadays, most approaches have focussed either on simple parametric subfamilies, or on semiparametric approaches combining univariate extreme value theory for the marginal distributions with non parametric estimation of dependence among marginals. Some example papers representing both approaches are Coles and Tawn (1991, 1994); Smith et al. (1997); de Haan and de Ronde (1998); Smith (2003). The problem resides in the multiple dependence among marginals. Indeed, in a parametric model the order of estimations based only on the dependence is 2^d, where d is the number of dimensions.

Recently, it has even been suggested that multivariate extreme value theory may not be a rich enough theory to encompass all the kinds of behaviour one would like to be able to handle, and alternative measures of tail dependence have been developed. The main references to this approach so far have been Ledford and Tawn (1996); Tawn (1988b); Poon et al. (2003); Maulik and Resnick (2004); Heffernan and Resnick (2005). In a more recent approach, due to the clear limitations of extreme value theory to higher dimensions, one theory based on a geometric approach has been developed by Balkema and Embrechts (2007). At the same time a semiprametric approach was proposed by Boldi and Davison (2007) which model the dependence among marginals by means of a mixture of Dirichlet densities.

In this chapter we work in a more wide context of extreme value theory. The infinite dimensional generalization of extreme value theory, which leads to max-stable processes, introduced by De Haan (1984). These processes have the potential to describe clustering behaviour. One of the most important features of max-stable processes is that they do not only model the dependence among the marginals, but also model the dependence across time.

In particular we concentrate on a class of max-stable processes introduced by Smith and Weissman (1996) to characterise the joint distribution of extremes in multivariate time series. They defined a class of max-stable processes, which they named Multivariate maxima of moving maxima process (M_4 processes for short)

$$M_4 \equiv X_{ij} = \max_{l \in \mathbb{N}} \max_{k \in \mathbb{Z}} a_{lkj} Z_{l,i-k}, \ i \in \mathbb{Z}, \quad j = 1, \ldots, d, \tag{5.1.1}$$

where a_{lkj} is a 3-dimensional matrix of non negative constants satisfying $\sum_l \sum_k a_{lkj} = 1$ for each j, and Z_{li} is a double sequence of iid unit Frèchet random variables. Note that there is no loss of generality in assuming unit Frèchet marginal distributions. We denote the array l as the *patterns* of the process and k as the *cluster size coefficient*.

The main focus of the paper by Smith and Weissman (1996) was to demonstrate that under fairly general conditions, extremal properties of a wide class of multivariate time series may be calculated by approximating the processes by one of M_4 form. Moreover, this approximation deals directly with the case of multivariate time series and not just of independent multivariate observations.

Another feature by which (5.1.1) is more directly useful for financial time series is that it represents the process in terms of an independent series of extreme values, whose large values among the Z_{li} determines the patterns or cluster in the extremes among the serie X_{ij}. The M_4 processes turn out to form a rich subclass of those of general multivariate stationary processes, mainly because stationary processes have the same multivariate extremal indexes as the M_4 processes have, i.e., the behaviour of the cluster of extremes in the multivariate framework is the same.

Unfortunately, there is an infinite number of parameters in the definition of M_4 processes, which poses challenges in statistical applications where workable models are preferred. Recently Zhang (2008) established sufficient conditions under which an M_4 process with infinite number of parameters may be approximated by an M_4 process with finite number of parameters. In this sense, the problem of modelling extremes of multivariate stationary processes can be stylized to the study of extremes of M_4 processes. However, the estimation of these class of processes is a challenging problem in itself, because of the fact that they suffer from degeneracies, i.e., the joint density of a set of random variables defined by (5.1.1) is typically singular with respect to Lebesgue measure and this causes problems for maximum likelihood techniques. Thus, other alternative of estimation has to be found.

The point is that, if we are capable to estimate a finite M_4 processes, we will be capable to approximate max-stable processes, i.e., multivariate extreme value processes, whose representation allows us to approximate multivariate time series with cluster at the extremes and heavy tail behaviour.

At the moment there are three feasible estimations for M_4 processes. Smith (2003) proposes to define candidate signature patterns on blocks of observations, and then uses a clustering method as k-means to indentify patter signatures. Zhang (2002) proposed a series of estimating procedures based on identifying signature patterns, on the bivariate distribution and weighted least squares estimations. Moreover, he showed consistency and asymptotic normality of the parameter estimators. Chamú (2005) proposes a state space representation of M_4 processes, where the state is an unobserved M_4 process, and the observed process is a nonlinear transformation on the state with small additive Gaussian noise to make the process nondegenerate. However, the complexity of the model makes it unsuitable for higher dimensions.

In this chapter we propose a nonparametric Bayesian approximation to the estimation of M_4 processes. The idea is to estimate the singularities in the multivariate density through an infinite mixture of Dirichlet processes of Gaussian distributions. These singularities have a cluster behaviour, which repeats infinitely, hence can be observed. The problem is to find out how many of these singularities exist. Until now, only approximations based on experience or arbitrary choices have been used. In this chapter we allow to deduce this number of singularities from the data. We assume that these singularities behave as an infinite multivariate Gaussian mixture, where a finite number is inferred from the available data observed.

This chapter is organized following. In Section 5.2 we outline relevant properties of the M_4 processes. In Section 5.3 we describe the main concepts in Dirichlet processes and Bayesian estimation apply to the estimation of M_4 processes. Section 5.6 shows some practical implementations of the models to estimate M_4 processes for different structures of the processes. Section 5.7 presents a application to the recent subprime crisis in the case of the German stock market. Conclusions and discussions are resumed in Section 5.8.

5.2. Properties of M_4 processes

The following introduction is necessary to understand the relation between M_4 processes and max-stable process. For more details in max-stable processes see Resnick (1987); De Haan (1984).

Consider a d-dimensional stochastic process Y_{ij} for $i = 1, \ldots, n$, $j = 1, \ldots d$. We are interested in the extremal properties of this process. Without loss of generality, we consider the case where all marginal distributions have unit Frèchet distribution.

DEFINITION 5.2.1. The process Y_{ij} is called max-stable if all finite dimensional distributions are max-stable, i.e., for any $n \geq 1$, $r \geq 1$

$$\mathbb{P}\left(Y_{ij} \leq r u_{ij} :\ 1 \leq i \leq n,\ 1 \leq j \leq d\right)^r = \mathbb{P}\left(Y_{ij} \leq u_{ij} :\ 1 \leq i \leq n,\ 1 \leq j \leq d\right).$$

Furthermore, a process X_{ij} for $i = 1, \ldots, n$, $j = 1, \ldots d$, is said to be in the domain of attraction of a max-stable process Y_{ij} if there exist normalising constants $a_{nij} > 0$, b_{nij} such that for any finite r

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_{ij} - b_{nij}}{a_{nij}} \leq r u_{ij} : 1 \leq i \leq n,\ 1 \leq j \leq j\right)^r = \mathbb{P}\left(Y_{ij} \leq u_{ij} :\ 1 \leq i \leq n,\ 1 \leq j \leq d\right). \tag{5.2.1}$$

Since we assume a priori that the process X_{ij} also has unit Frèchet, then we may take $a_{nij} = n$, $b_{nij} = 0$.

We can apply this definition to show that the M_4 process defined in (5.1.1) is in fact a max-stable process

$$
\begin{aligned}
M_4(\boldsymbol{u}, n) &= \mathbb{P}\left(X_{ij} \le u_{ij},\ 1 \le i \le n,\ 1 \le j \le d\right) \\
&= \mathbb{P}\left(Z_{l,i-k} \le \frac{u_{ij}}{a_{lkj}},\ l \in \mathbb{N},\ k \in \mathbb{Z},\ 1 \le i \le n,\ 1 \le j \le d\right) \\
&= \mathbb{P}\left(Z_{lm} \le \min_{1-m \le k \le n-m} \min_{1 \le j \le d} \frac{a_{lkj}}{u_{m+k,j}},\ l \in \mathbb{N},\ m \in \mathbb{Z}\right) \qquad (5.2.2) \\
&= \exp\left(-\sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} \max_{1-m \le k \le n-m} \max_{1 \le j \le d} \frac{a_{lkj}}{u_{m+k,j}},\ l \in \mathbb{N}\right) \\
&= \prod_{l \in \mathbb{N}} \prod_{m \in \mathbb{Z}} \exp\left(- \max_{1-m \le k \le n-m} \max_{1 \le j \le d} \frac{a_{lkj}}{u_{m+k,j}},\ l \in \mathbb{N}\right)
\end{aligned}
$$

and from (5.2.2), it is easy to see that

$$
\mathbb{P}^r\left(X_{ij} \le r u_{ij},\ 1 \le i \le n,\ 1 \le j \le d\right) = \mathbb{P}\left(X_{ij} \le u_{ij},\ 1 \le i \le n,\ 1 \le j \le d\right),
$$

which tells that X_{ij} are max-stable.

Smith and Weissman (1996) characterized the conditions under which the multivariate extremal index from a stationary time series could be calculated from a max-stable process with the same limiting distributions for any finite dimensional multivariate extremes. Furthermore, they show that any max-stable process in d-dimensions could be approximated with arbitrary accuracy by one of M_4 form through a direct generalisation of the result of Deheuvels (1983) for one dimensional case. To derive this result we need first some mixing condition.

Theorem 5.2.2. *(Mixing condition) Let $\tau = \{\tau_1, \ldots, \tau_d\}$, $\tau_j \in [0, \infty]$, $j = 1, \ldots, d$. For a given sequence of thresholds $\mathbf{u}_n = \{u_1, \ldots, u_d\}$. Since Z_{lk} is unit Frèchet we can take $u_{nj} = n/\tau_j$ such that $n\left(1 - F_j(u_{nj})\right) \to \tau_j$. For $1 \le m \le k \le n$, let $\mathcal{B}_1^k(\mathbf{u}_n)$ denote the σ-field generated by the events $\{X_{ij} \le u_{ij},\ m \le i \le k\}$, and for each integer t let*

$$
\alpha_{nt} = \sup\left\{\left|\mathbb{P}\left(A \cap B\right) - \mathbb{P}(A)\mathbb{P}(B)\right| :\ A \in \mathcal{B}_1^k(\mathbf{u}_n),\ B \in \mathcal{B}_{k+t}^n(\mathbf{u}_n)\right\},
$$

where the supremum is taken over $1 \le k \le n - t$ and two respective σ- fields. Following Nandagopalan (1994), the condition $\triangle(u_n)$ is said to hold if there exists a sequence $\{t_n\}_{n \ge 1}$ such that $t_n \to \infty$, $t_n/n \to 0$, $\alpha_{n,t_n} \to 0$ as $n \to \infty$. Assume $\triangle(u_n)$ holds with respect to some sequence $\{t_n\}_{n \ge 1}$, and define a sequence $\{k_n\}_{n \ge 1}$ such as $k_n \to \infty$, $k_n t_n/n \to 0$, $k_n \alpha_{n,t_n} \to 0$ as $n \to \infty$.

Definition 5.2.3. We assume that, with $r_n = \lfloor n/k_n \rfloor$,

$$
0 = \lim_{r \to \infty} \lim_{n \to \infty} \sum_{i=r}^{r_n} \sum_{j=1}^{d} \mathbb{P}\left(X_{ij} > u_{nj} \mid \max_j \left(\frac{X_{1j}}{u_{nj}}\right) > 1\right). \qquad (5.2.3)
$$

We now state the main Theorem of Smith and Weissman (1996).

THEOREM 5.2.4. *(Theorem 2.3 in Smith and Weissman (1996)). Suppose the processes* $\mathbf{X}$ *and* $\mathbf{Y}$ *are each stationary with unit Frèchet distribution and equation (5.2.1) holds. For a given* $\tau = \{\tau_1, \ldots, \tau_d\}$, *and* $u_{nj} = n/\tau_j$ *suppose* $\triangle(u_n)$ *and (5.2.3) hold for both* $\mathbf{X}$ *and* $\mathbf{Y}$. *Then the two processes* $\mathbf{X}$ *and* $\mathbf{Y}$ *have the same multivariate extremal index* $\theta(\tau)$.

PROOF. See Smith and Weissman (1996). □

The proof is based on the fact that all max-stable process can be rewritten in terms of a M_4 process and another process, which is maximum over periodic deterministic sequences, and therefore a perfectly predictable process. For this reason, it seems reasonable keep only the M_4 process, with the fact that the deterministic process cannot occur in most applications.

Notice that we can rewrite model (5.2.2) for positive integer n and thresholds $\boldsymbol{u} = \{u_1, \ldots, u_d\}$, as

$$\begin{aligned} M_4(\boldsymbol{u}, n) &= \exp\left(-\sum_{l\in\mathbb{N}}\sum_{k\in\mathbb{Z}} \max_{i=1,\ldots,n} \max_{j=1,\ldots d} \frac{a_{l,i-k,j}}{u_j}\right) \\ &= \exp\left(-W_4(\boldsymbol{u}, n)\right). \end{aligned} \tag{5.2.4}$$

Under this notation we can define the following lemma, which will be of great use in the study of the temporal dependence between extremes of a M_4 process.

LEMMA 5.2.5. *Let* $\boldsymbol{u} = \{u_1, \ldots, u_d\}$, *the representation* $W_4(\boldsymbol{u}, n)$ *in (5.2.4) satisfies*

$$\begin{aligned} \lim_{n\to\infty}\left(W_4(\boldsymbol{u}, n) - W_4(\boldsymbol{u}, n-1)\right) &= \lim_{n\to\infty} W_4(\boldsymbol{u}n, n) = \sum_{l\in\mathbb{N}} \max_{r\in\mathbb{Z}} \max_{j=1,\ldots d} \frac{a_{lrj}}{u_j} \\ &= W_4'(\boldsymbol{u}). \end{aligned}$$

It is easy to see that for $\boldsymbol{u} = \{u_1, \ldots, u_d\}$ Lemma 5.2.5 provides a direct estimation of the multivariate extremal index:

$$\theta(\boldsymbol{u}) = \frac{W_4'(\boldsymbol{u})}{W_4(\boldsymbol{u},1)} = \frac{\sum_{l\in\mathbb{N}} \max_{r\in\mathbb{Z}} \max_{j=1,\ldots d} a_{lrj}/u_j}{\sum_{l\in\mathbb{N}} \sum_{r\in\mathbb{Z}} \max_{j=1,\ldots d} a_{lrj}/u_j}. \tag{5.2.5}$$

The last result in equation (5.2.5) is due to Smith and Weissman (1996). Similarly we can derive two interesting limits for the multivariate extremal index when the M_4 processes reach completely dependence and independence. By equation (5.2.2) we can observe that

$$\begin{aligned} M_4 &= \prod_{l\in\mathbb{N}}\prod_{m\in\mathbb{Z}} \exp\left(-\max_{1-m\le k\le r-m} \max_{1\le j\le d} \frac{a_{lkj}}{u_{m+k,j}}\right) \\ &= \prod_{l\in\mathbb{N}}\prod_{m\in\mathbb{Z}} C_{M_4}\left(\exp\left(-\max_{1-m\le k\le r-m} \frac{a_{lk1}}{u_{m+k,1}}\right)\cdots \exp\left(-\max_{1-m\le k\le r-m} \frac{a_{lkd}}{u_{m+k,d}}\right)\right) \end{aligned}$$

where C_{M_4} is the dependence function, which describes the dependence among the marginals. In the standard case M_4 the dependence function is $C_{M_4}(x_1, \ldots, x_d) = \min_{1\le j\le d}(x_1, \ldots x_d)$. Thus, the standard model is always asymptotic dependent. The complete independence case arises when we take $C_{M_4}(x_1,\ldots,x_d) = \prod_{j=1}^{d}(x_1,\ldots x_d)$, which reduces to

$$M_4^{(ind)} = \prod_{l\in\mathbb{N}}\prod_{m\in\mathbb{Z}}\prod_{1\le j\le d}\exp\left(-\max_{1-m\le k\le r-m}\frac{a_{lkj}}{u_{m+k,j}}\right).$$

In this case the multivariate extremal index for a threshold $\boldsymbol{u} = \{u_1,\ldots,u_d\}$ takes he form:

$$\theta(\boldsymbol{u}) = \frac{\sum_{l\in\mathbb{N}}\max_{r\in\mathbb{Z}}\max_{j=1,\ldots d} a_{lrj}/u_j}{\sum_{l\in\mathbb{N}}\sum_{r\in\mathbb{Z}}\sum_{j=1}^{d} a_{lrj}/u_j},$$

which can be derived as in (5.2.5). Of course, other kinds of dependence function C_{M_4} could be used, however, the difficulty to estimate the processes would cause these classe of dependence lack interest.

Tail dependece is an important property when working with extreme value models. For instance, the M_4 has always a postive tail dependence, which means that it is not a suitable model when asymptotic independence could be present in the data.

Proposition 5.2.6. *(Tail dependence of M_4 processes). Consider a bivariate M_4 process. Then, the tail dependence is defined as*

$$\lambda = 2 - \sum_{l\in\mathbb{N}}\sum_{k\in\mathbb{Z}}\max_{1\le j\le 2} a_{lkj}.$$

Other probabilistic properties of M4 processes can be found in Zhang and Smith (2004). An extension to the asymptotic independent case can be found in Heffernan et al. (2007).

Now, we given some idea of how we can estimate the M_4 process with finite number of parameters based on the discontinuity of the multivariate density.

5.2.1. On the estimation of M_4 processes. The statistical parameter estimation of the infinite dimensional parameter model (5.1.1) is almost impossible. Zhang (2008) establishes sufficient conditions under which an M_4 process with infinite number of parameters may be approximated by an M_4 process with finite number of parameters. These conditions provide us with the probabilistic basis required for application to real data. A finite order M_4 process can be written in the following form:

$$M_4(\mathcal{L},\mathcal{K},d) \equiv X_{ij} = \max_{l\in\mathcal{L}}\max_{k\in\mathcal{K}} a_{lkj}Z_{l,i-k},\ i\in\mathbb{Z},\quad j=1,\ldots,d. \tag{5.2.6}$$

To motivate the method that we will provide to estimate a wide class of M_4 process, suppose for the moment that we have observed a 1-dimesional $M_4(\mathcal{L},\mathcal{K},d) = M_4(1,\mathcal{K},1)$ process with $\mathcal{L}=1$, where k indicates a infinite number (countable) of observations before and after the observation i that we have observed, i.e., the cluster size is exactly $\mathcal{K} = 2k+1$. We will denote the set $\mathcal{K} = \{-k,\ldots,0,\ldots,k\}$ as the *cluster size.* Let us define the process

$$M_4(1,\mathcal{K},1) \equiv X_i = \max_{k\in\mathcal{K}} a_k Z_{i+k},\ i\in\mathbb{Z}, \tag{5.2.7}$$

and that for some i^* the value of Z_{i^*} is much larger than its neighbors $Z_{i*-k},\ldots,Z_{i*-1},Z_{i*+1},Z_{i*+k}$, such that $X_{i*+k} = a_k Z_{i*}$. Note that a *signature pattern* for this process is determined by

$$\frac{X_{i*+k_1}}{X_{i*+k_2}} = \frac{a_{k_1}}{a_{k_2}}, \tag{5.2.8}$$

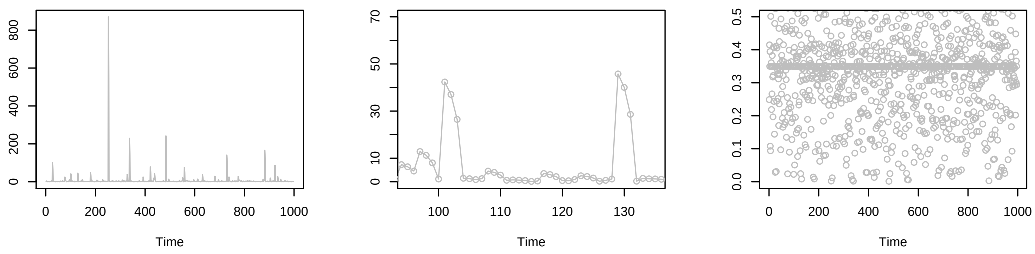

FIGURE 5.2.1. Simulation of 1.000 observations of $M_4\,(1,\mathcal{K},1)$ process proposed in Example 5.2.7. (Left) Plot of the process. (Middle) Plot of the process partialy drawn from the whole simulated data showing two clusters. (Right) Plot of the ratios $X_i/X_{i-1}+X_i+X_{i+1}$ of 1.000 observations.

where k_1,$k_2 \in \mathcal{K}$. Since that $\sum_{\mathcal{K}} a_k = 1$ and solving for some a_k in (5.2.8) we obtain

$$a_k = \frac{X_{i*+k}}{\sum_{\mathcal{K}} X_{i*+k}}. \tag{5.2.9}$$

EXAMPLE 5.2.7. Define a $M_4\,(1,\mathcal{K},1)$ process with $a_k = \{a_{-1}, a_0, a_1\} = \{0.25, 0.35, 0.40\}$. Observe that the coefficient a_i can be obtained by the singularities of the ratios $X_i/X_{i-1} + X_i + X_{i+1}$,$X_{i-1}/X_{i-1} + X_i + X_{i+1}$, $X_{i+1}/X_{i-1} + X_i + X_{i+1}$. This approach is displayed numerically in Figure 5.2.1.

Smith (2003) showed that these signatures will hold infinitely often for high values of Z_{i^*}, if we can observe the process for a long period of time. This process will create a deterministic pattern, which yields to the determination of the coefficients a_k and we will say that the process is *identifiable.*

Following the same argument, we can extend this result to higher array of patterns $\mathcal{L} \geq 1$ and number of dimensions d. Let

$$M_4\,(\mathcal{L},\mathcal{K},d) \equiv X_{ij} = \max_{l\in\mathcal{L}} \max_{k\in\mathcal{K}} a_{l,k,j} Z_{l,i+k},\ i = 1,\ldots,n,\ j = 1,\ldots,d.$$

Smith (2003, equation 3.5) showed that for M_4 processes the relative frequency of the l-th signature pattern is proportional to $a_l^{\Sigma} = \sum_{\mathcal{K}} \max_d a_{lkj}$ for all $l \in \mathcal{L}$. This is the key feature that is used for identifying the parameters in a M_4 process.

Notice that this sum reflects the proportion of time, in which the process is driven by the l-th pattern in the d-dimensional process. Without loss of generality we define the a *multivariate local maximum* as

$$X_{i^*,j^*} \geq \max\{\boldsymbol{u}, \boldsymbol{X}_{i+k}\}, \tag{5.2.10}$$

where $\boldsymbol{X}_i = \{X_{i1},\ldots,X_{id}\}$ and $\boldsymbol{u}$ is a high d-variate high threshold level. Define the set $\mathcal{D} = \{(k,j):\ k\in\mathcal{K},\ j = 1,\ldots,d\}$, then (5.2.10) implies $a_{l^*j^*} = a_{l^*,0,j^*} = \max_{(k,j)\in\mathcal{D}} a_{l^*,k,j^*}$.

Assume that there exist a Z_{l^*,i^*} much larger than its neighbors such that for $(k,j)\in\mathcal{D}$

$$X_{i^*+k,j} = a_{l^*,k,j} Z_{l^*,i^*},$$

Thus, the l^*-th signature pattern is for example a $(2k+1)\times d$ matrix represented by

$$\Psi_{l^*,i^*+k,j} = \frac{X_{i^*+k,j}}{X_{i^*,j^*}} = \frac{a_{l^*,k,j}}{a_{l^*,j^*}} \quad \forall_{k,j} \in \mathcal{D}. \tag{5.2.11}$$

Let us define

$$X_i^{\Sigma} = \sum_{\mathcal{K}} \max_j X_{i+k,j}, \quad \text{and} \quad a_l^{\Sigma} = \sum_{\mathcal{K}} \max_j \alpha_{l,k,j}. \tag{5.2.12}$$

Observe that

$$\sum_{\mathcal{K}} \Psi_{l^*,i^*+k,j} = \frac{X_{i^*}^{\Sigma}}{X_{i^*j^*}} = \frac{a_{l^*}^{\Sigma}}{a_{l^*j^*}},$$

which implies that

$$a_{l^*j^*} = a_{l*}^{\Sigma}\frac{X_{i^*j^*}}{X_{i^*}^{\Sigma}}. \tag{5.2.13}$$

Substituting (5.2.13) into (5.2.11) we get

$$y_{l^*,p} = \frac{\boldsymbol{a}_{l^*,k}}{a_{l^*}^{\Sigma}} = \frac{\boldsymbol{X}_{i^*+k}}{X_{i^*}^{\Sigma}} \quad \forall_{i,j} \in \mathcal{D}, \tag{5.2.14}$$

where $y_{l^*,p}$ is a $1\times(d\,(2k+1))$ vector with

$$p = \{1,\ldots,P\} = \{(-k,1),(-k+1,1),\ldots,(0,1),\ldots,(k-1,1),(k,1),\ldots,(k,d)\}.$$

Equivalently,

$$a_{l^*,k,j} = a_{l*}^{\Sigma} y_{l^*,p} = a_{l*}^{\Sigma}\frac{X_{i^*+k,j}}{X_{i^*}^{\Sigma}} \quad \forall_{i,j} \in \mathcal{D}. \tag{5.2.15}$$

The last two equations give us the key to estimate M_4 processes. In a d- dimensional process for a given threshold $\boldsymbol{u}$, The idea is to calculate (5.2.14) whenever a multivariate local maximum is observed. If we can classify each signature pattern $y_{l^*,p}$ in $\mathcal{L}$ signature patterns, an estimator of a_{l*}^{Σ} can be obtained up to proportionaty by $n_{l^*}/n_{\mathcal{L}}$, where n_{l^*} is the number of signatures associated to the pattern l^* present in the sample and $n_{\mathcal{L}}$ is the total of signatures.

Notice that we transform the process X_{ij} with matrix size $n\times d$ to one y_{lp} with matrix size $\frac{N_u}{\mathcal{K}}\times P$, where N_u is the number of exceedances of at least one observation in the d-dimensional process X_{ij} over the threshold $\boldsymbol{u}$, i.e., $X_{ij} \nleq u_j$ for $j=1,\ldots,d$, and $P = d\cdot\mathcal{K}$. Thus, the problem is now a P-dimensional problem. We will call the process y_{lp} as the *standardized process of signature patterns*. We formalized the results in the following proposition.

Proposition 5.2.8. *Let* $X_{ij} = \max_{l\in\mathcal{L}}\max_{\mathcal{K}} a_{l,k,j}Z_{l,i+k}$, $i\in\mathbb{Z}$, $j=1,\ldots,d$ *be a* d-*dimensional* M_4 *process with cluster size coefficient* $\mathcal{K}$ *and array patterns* $\mathcal{L}$. *Further, let the Frèchet random variable* Z_{l^*,i^*} *be much larger than its neighbors, such that* $X_{i*+k,j} = a_{l^*,i*+k,j}Z_{l^*,i*}$. *Then,* $a_{l*}^{\Sigma} = \sum_{\mathcal{K}}\max_j a_{l^*,k,j}$ *reflects the proportion of clusters with the characteristic of the* l^*-*th array pattern that should appear in the* d-*dimensional process, and the coefficients of the process are identifiable by* $a_{l^*,k,j} = a_{l*}^{\Sigma}\frac{X_{i^*+k,j}}{X_{i^*}^{\Sigma}} \quad \forall_{i,j}\in\mathcal{D}$.

Proof. The demonstration follows. □

Proposition 5.2.8 has two important implications:

- First, each of the $\mathcal{L}$ signature patterns will occur infinitely often. This means that the joint density of M_4 processes contains singularities because of the presence of

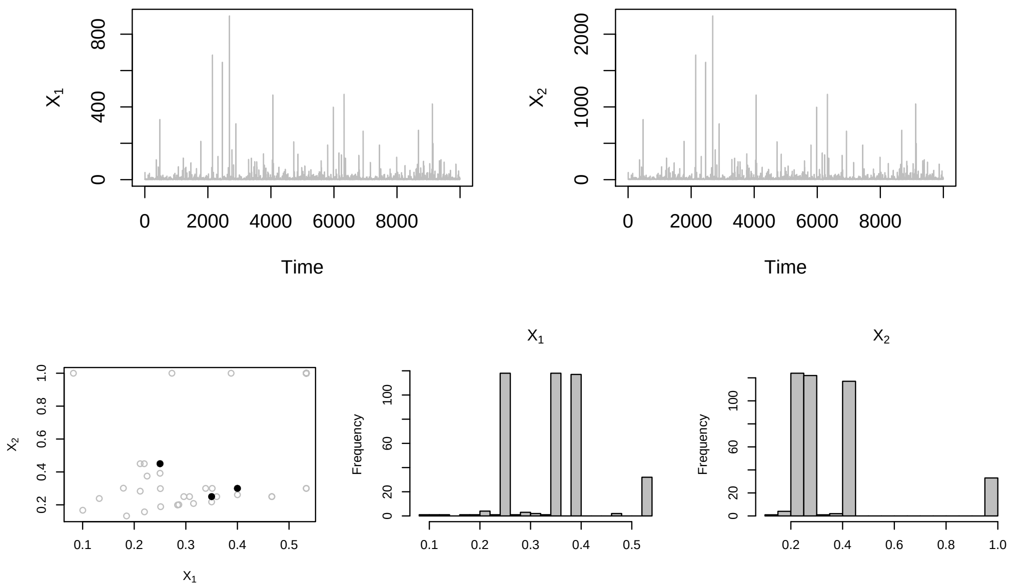

FIGURE 5.2.2. (Top panel) Simulation of a $M_4\,(1,3,2)$ process with 10.000 observations proposed in Example 5.2.9. (Left bottom) Plot of the ratios $\left(\frac{X_{i*-1,1}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*,1}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*+1,1}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*-1,2}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*,2}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*+1,2}}{\sum_{\mathcal{K}} X_{i*+k,1}}\right)$ for $X_{i,j}$ over its 0.95-th quantile . (Middle-bottom) Histogram of the sigularity clusters for $X_{i,1}$. (Rigth-bottom) Histogram of the sigularity clusters for $X_{i,2}$.

these deterministic patterns. Therefore, it is not possible to apply the method of maximum likelihood to estimate the parameters of the model.

- Second, the estimation of the standardized signature pattern y_{lp} for $l \geq 1$ and $j = 1, \ldots, d$ is equivalent to estimate the number of patterns $\mathcal{L}$ in the multivariate process.

Smith (2003) proposes to define candidate signature patterns on blocks of extreme observations over a high threshold u_j for each marginal. He calculates the ratios $\frac{X_{i*+k}}{\sum_K X_{i*+k}}$, and then uses a clustering algorithm to identify the number of observations n_{l^*} due to the pattern l^* to approximate $a_{l^*}^{\Sigma}$. Posteriory, with equation (5.2.15) estimate the coefficients a_{l^*i*+k} given $\mathcal{K}$.

EXAMPLE 5.2.9. Define a $M_4\,(1,3,2)$ process with $\boldsymbol{a}_{k,1} = \{a_{-1,1}, a_{0,1}, a_{1,1}\} = \{0.25, 0.35, 0.40\}$ and $\boldsymbol{a}_{k,2} = \{a_{-1,2}, a_{0,2}, a_{1,2}\} = \{0.45, 0.25, 0.30\}$. Observe that the coefficient $(\boldsymbol{a}_{k,1}, \boldsymbol{a}_{k,2})$ can be obtained by the singularities of the ratios $\left(\frac{X_{i*-1,1}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*,1}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*+1,1}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*-1,2}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*,2}}{\sum_{\mathcal{K}} X_{i*+k,1}}, \frac{X_{i*+1,2}}{\sum_{\mathcal{K}} X_{i*+k,1}}\right)$ for $X_{i,j}$ over its 0.95-th quantile. Note that the estimation produces a 6-dimensional vector. This approach is displayed jointly in Figure 5.2.2.

In practice it is unrealistic to assume that we can observe the exact $\mathcal{L}$ patterns and the cluster size coefficient $\mathcal{K}$, which are characteristic of M_4 processes. Further, we would not expect to observe an exact M_4 process, but a M_4 process with some kind of noise. While this noise has lighter tails than a Frèchet random variable, the pure M_4 process is still estimable. For simplicity we show this result for a bivariate Gaussian noise.

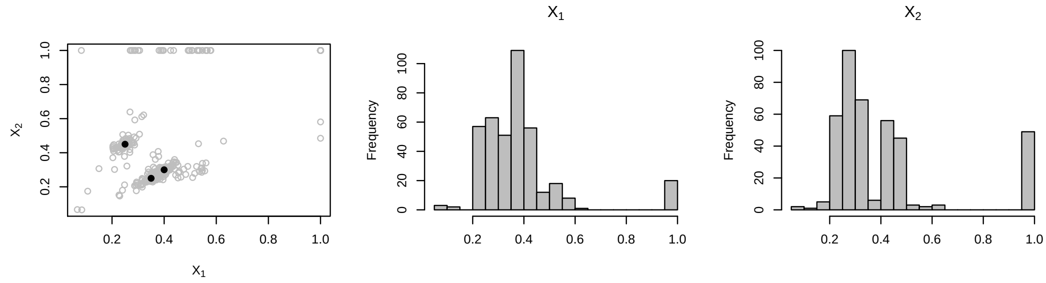

FIGURE 5.2.3. (Top panel) Simulation of a $M_4\,(1,1,2)$ process with a bivariate Gaussian noise N proposed in Example 5.2.11. The Sample size is 10000 . (Left) Plot of the ratios $\left(\frac{X_{i+p,1}}{X_{i-1,1}+X_{i,1}+X_{i+1,1}}, \frac{X_{i+p,2}}{X_{i-1,2}+X_{i,2}+X_{i+1,2}}\right)$ for $X_{i,j}$ over its 0.95-th quantile. (Middle) Histogram of the sigularity clusters for $X_{i,1}$. (Rigth) Histogram of the sigularity clusters for $X_{i,2}$.

PROPOSITION 5.2.10. *Let us observe a bivariate M_4 process plus a bivariate noise* $\boldsymbol{N} = \left(N_1\sqrt{1-\rho^2}+\rho N_2, N_2\right)$, *where $N_1$and N_2 are two Normal random variables with correlation* ρ.

$$M_4\,(\mathcal{L},\mathcal{K},2) \equiv X_{ij} = \max_{l\in\mathcal{L}}\max_{k\in\mathcal{K}} a_{l,k,j}Z_{l,i+k} \quad + \boldsymbol{N},\ i\in\mathbb{Z},\ j=1,2.$$

For a fixed marginal j and a large clustered observations, we have $a_{l^*k} = a^{\Sigma}_{l^*}E\left(\frac{X_{i*+k}}{\sum_K X_{i*+k}}\right)$

In resume, we can estimate a M_4, even if a light tailed noise in the process is present. Notice that the result does not depend on the kind of noise, as long as the tails are lighter than the tails of a M_4 process.

EXAMPLE 5.2.11. (Cont.) Define a $M_4\,(1,3,2)$ process with $\boldsymbol{a}_{k,1} = \{a_{-1,1}, a_{0,1}, a_{1,1}\} = \{0.25, 0.35, 0.40\}$ and $\boldsymbol{a}_{k,2} = \{a_{-1,2}, a_{0,2}, a_{1,2}\} = \{0.45, 0.25, 0.30\}$ as in Example 5.2.9 plus a bivariate Gaussian noise $\boldsymbol{N}$ with mean zero and covariance matrix $\Sigma = \begin{pmatrix} 20 & 15 \\ 15 & 20 \end{pmatrix}$. By Proposition 5.2.10, observe that the coefficient $(\boldsymbol{a}_{k,1}, \boldsymbol{a}_{k,2})$ can be obtained by the singularities of the ratios $\left(\frac{X_{i*-1,1}}{\sum_K X_{i*+k,1}}, \frac{X_{i*,1}}{\sum_K X_{i*+k,1}}, \frac{X_{i*+1,1}}{\sum_K X_{i*+k,1}}, \frac{X_{i*-1,2}}{\sum_K X_{i*+k,1}}, \frac{X_{i*,2}}{\sum_K X_{i*+k,1}}, \frac{X_{i*+1,2}}{\sum_K X_{i*+k,1}}\right)$ even under the presence of a noise process $\boldsymbol{N}$. Notice that this defines a 6 dimensional matrix. The ratios are obtained for $X_{i,j}$ over its 0.95-th quantile. This approach is displayed jointly in Figure 5.2.3 for each pair of dimensions.

The results of Smith and Weissman (1996) allow us to characterize the extremal behaviour of a multivariate stationary time series in terms of a limiting max-stable process. However, there has been little work on the statistical modelling of max-stable processes.

An important issue that must be addressed in M_4 processes is the question of how wide the cluster size coeffcient $\mathcal{K}$, (the serial dependence) and patterns $\mathcal{L}$ (the different forms or shapes of the clusters) are to use. Bayesian statistics and model based approaches can provide elegant solutions to model selection questions of this kind. By now, we are still supposing that the cluster size coefficient $\mathcal{K}$ is known, and we leave its discussion for the empirical section. However, there is not too much to do in relation to $\mathcal{K}$, therefore it should not be larger than the major cluster size observed for the sample X_{ij} over a multivariate threshold $\boldsymbol{u}$.

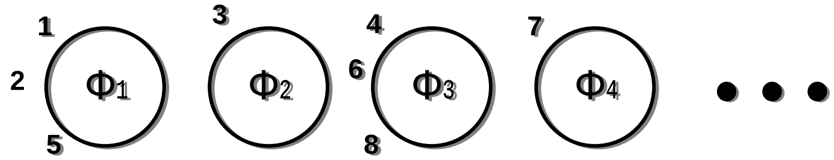

FIGURE 5.3.1. The Chinese restaurant process mixture with l' equal to four tables and eight customers. In this example, the 1st, 3rd, 4th, and 7th customers all sat at an empty table, whereas the 2nd, 5th, 6th, and 8th sat at accupied tables. The 9th customer will sit at table 1, 2, 3, or 4 with probabilities $\frac{3}{8+\alpha}, \frac{1}{8+\alpha}, \frac{3}{8+\alpha}$, and $\frac{1}{8+\alpha}$ respectively, or will sit at a new table with probability $\frac{\alpha}{8+\alpha}$.

Within a Bayesian framework, all assumptions are presented in terms of priors and the choice of a likelihood function. In this framework each signature $\boldsymbol{y_l} = \{y_1, \ldots, y_p\}$, as it was defined in (5.2.14), represents a P-multidimensional vector of measurements. Further, the probability distribution for each variable $\boldsymbol{y_l}$ is assumed to be a multivariate Gaussian distribution. We describe an approach to the problem of automatically determine the number patterns $\mathcal{L}$ based on the theory of infinite Gaussian mixtures or Dirichlet process mixtures. This theory is based on the observation that the mathematical limit of an infinite number of components in an ordinary finite mixture model (i.e. the patterns $\mathcal{L}$ in the d-dimensional model) corresponds to a Dirichlet process prior. In an infinite Gaussian mixture model there is no need to make arbitrary choices about how many patterns $\mathcal{L}$ there are in the process. The major advantage is that, although in theory the infinite mixture model has an infinite number of parameters, it is possible to do exact inference in these infinite mixture models efficiently using Markov chain Monte Carlo (MCMC) methodology.

5.3. Nonparametric Bayesian analyses of mixture distributions

Dirichlet process (DPM) mixture models are the cornerstone of nonparametric Bayesian statistics. The development of Monte Carlo Markov chain (MCMC) sampling methods for DPMs has enabled the application of nonparametric Bayesian methods to a variety of practical data analysis problems.

DPM models have attracted much attention recently, because they are applicable even if the number of mixtures is not known, as in our application. The properties of the DP enable the model to uncover clusters and determine the number of clusters. An intuitive approach to how a DP works, is to consider a sampling scheme known as the Chinese Restaurant Process.

Imagine a restaurant with countable infinitely many tables, labelled $1, 2, \ldots, \mathcal{L} = \infty$. Customers walk in and sit down at one table. The tables are chosen according to the following random process. The first customer always chooses the first table and orders a dish. The second customer enters and decides either to sit at the first table with a probability $1/1 + \alpha$

or a new table with probability $\alpha/1+\alpha$. When sitting at a new table the customer orders a new dish. This process continues for each new customer. Thus, the l-th customer chooses a new unoccupied table with probability $\alpha/\mathcal{L}-1+\alpha$, and an occupied table with probability $n_l/\mathcal{L}-1+\alpha$, where n_l is the number of people sitting at the table l. In the above, α is a scalar parameter of the process. One might acknowledge that the above does define a probability distribution. Let us denote by $l^{'}$ the number of different tables occupied after $\mathcal{L}$ customers have walked in. Then $1 \leq l^{'} \leq \mathcal{L}$ and it follows the above description that precisely $l^{'}$ tables are occupied. Notice that popular tables become less and less likely to sit down at a new table. In this representation the dishes correspond to probability density functions, and the process of ordering a dish l corresponds to drawing the parameters ϕ_l to a probability density function, as for example a Gaussian from a prior distribution G over those parameters. The process of a customer l choosing a table c_l corresponds to choosing a distribution ϕ_{c_l} from which to draw an observation y_l. Since the structure of the process, is that customers tend to sit at tables with many other customers producing the cluster behaviour, thought it has an infinite number of mixture components to choose from. Furthermore, the expected number of occupied tables for $\mathcal{L}$ customers grows logarithmically. In particular $\mathbb{E}\left[l^{'} \mid \mathcal{L}\right] = \sum_{l=1}^{\mathcal{L}} \frac{\alpha}{\alpha+l-1} \in \mathcal{O}\left(\alpha \log \mathcal{L}\right)$.

This slow growth of the number of clusters makes sense because of the rich-gets-richer phenomenon: we expect there to be large clusters thus the number of clusters $l^{'}$ to be far smaller than the number of observations $\mathcal{L}$. Notice that α controls the number of clusters in a direct manner, with larger implying a larger number of clusters a priori. This intuition will help in the application of DPs to mixture models. The model allows a priori infinite number of patterns $l^{'}$.

These patterns l' will represent in our case the number of observed of different signatures l. At the beginning we will suppose that the true number of signatures is $\mathcal{L} \rightarrow \infty$ in a mode that we will examplain later, which is the natural and more flexible definition of M_4 processes.

5.3.1. Dirichlet processes. The Dirichlet distribution forms our first step toward understanding the DP model. The vector $\{y_1, \ldots, y_n\}$ has a Dirichlet distribution with parameter $\{\alpha_1, \ldots, \alpha_n\}$, if its density function is

$$f\left(y_1, \ldots, y_n \mid \alpha_1, \ldots, \alpha_n\right) = \frac{\Gamma\left(\sum_{i=1}^{n} \alpha_i\right)}{\prod_{i=1}^{n} \Gamma\left(\alpha_i\right)} \prod_{i=1}^{n} y_i^{\alpha_i - 1} \qquad \text{for } y_i \geq 0,\ \sum_{i=1}^{n} y_i = 1,$$

where $\alpha_i > 0$ for $i = 1, \ldots, n$. The class of Dirichlet distributions is known to Bayesians as the conjugate prior for the parameters of a multinomial distribution. Some examples are displayed in Figure 5.3.2 for different values of α.

A class of prior probability distributions $f\left(y\right)$ is said to be conjugated to a class of likelihood functions $f\left(y \mid \alpha\right)$ if the resulting posterior distributions $f\left(\alpha \mid y\right)$ are in the same family as $f\left(y\right)$. The main properties of the class of Dirichlet distributions are

(1) If $\{y_1, \ldots, y_n\}$ has a Dirichlet distribution with parameters $\{\alpha_1, \ldots, \alpha_n\}$ and $r_1, \ldots, r_n$ are integers such that $0 < r_1 < \cdots < r_n = n$, then

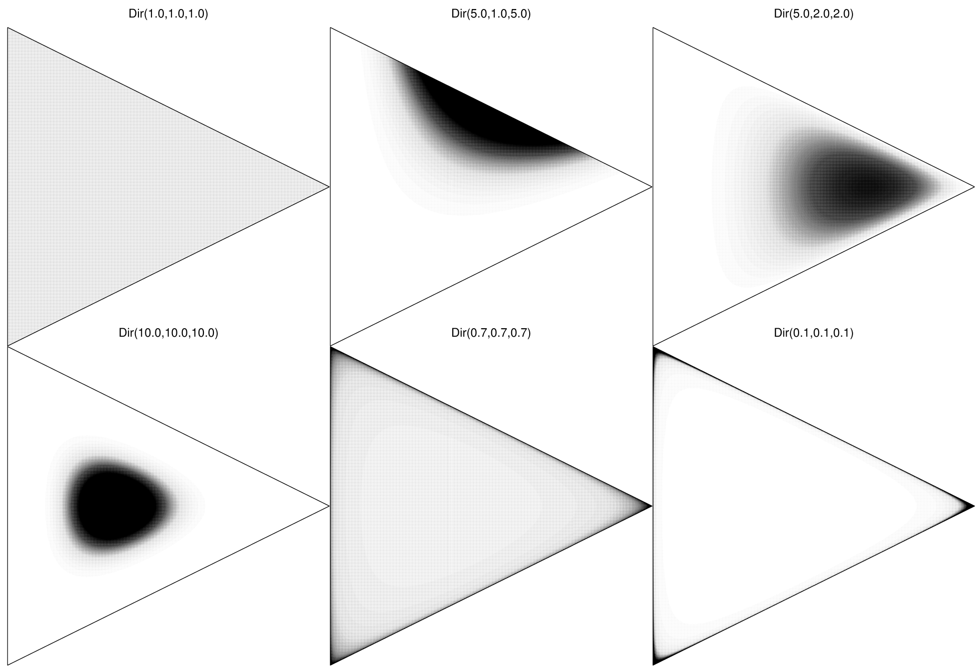

FIGURE 5.3.2. Some examples of trivariate Dirichlet distributions.

$\left\{\sum_{i=1}^{r_1} y_i, \sum_{i=r_1+1}^{r_2} y_i, \ldots, \sum_{i=r_{n-1}+1}^{r_n} y_i\right\}$ has a Dirichlet distribution with parameter $\left\{\sum_{i=1}^{r_1} \alpha_i, \sum_{i=r_1+1}^{r_2} \alpha_i, \ldots, \sum_{i=r_{n-1}+1}^{r_n} \alpha_i\right\}$.

(2) If $\{y_1, \ldots, y_n\}$ has a Dirichlet distribution with parameters $\{\alpha_1, \ldots, \alpha_n\}$, then

$$\begin{aligned}\mathbb{E}[y_i] &= \alpha_i/\alpha, \\ \mathbb{E}\left[y_i^2\right] &= \alpha_i(\alpha_i+1)/(\alpha(\alpha+1)), \\ \mathbb{E}[y_i y_j] &= \alpha_i\alpha_j/(\alpha(\alpha+1)) \quad \text{for } i \neq j,\end{aligned}$$

where $\alpha = \sum_{i=1}^{n} \alpha_i$.

(3) If $\{y_1, \ldots, y_n\}$ has a Dirichlet distribution with parameters $\{\alpha_1, \ldots, \alpha_n\}$ and if

$$\mathbb{P}(X = j \mid y_1, \ldots, y_n) = y_j \qquad \text{for } j = 1, \ldots, n,$$

then the posteriori distribution of $\{y_1, \ldots, y_n\}$ given $X = j$ is a Dirichlet distribution with parameter $(\alpha_1^*, \ldots, \alpha_n^*)$ where

$$\alpha_i^* = \begin{cases} \alpha_i & if\ i \neq j \\ \alpha_i + 1 & if\ i = j. \end{cases}$$

Now we are able to introduce the DP. The DP is simply an extension of the Dirichlet distribution to continuous spaces, and therefore a measure on measures. Formally, the DP is a stochastic process whose sample paths are probability measures with probability one. Stochastic processes are distributions over function spaces, with sample paths being random functions drawn from the distribution. In the case of the DP, it is a distribution over probability measures, which are functions with certain special properties which allow them to be interpreted as distributions over some probability space. Thus, draws from a DP can be interpreted as random distributions. For a distribution over probability measures to be a DP, its marginal distributions have to take on a specific form which we shall give below.

DEFINITION 5.3.1. We say G is DP distributed with base distribution G_0 and concentration parameter α, written $G \sim DP(\alpha, G_0)$, if

$$(G(A_1), \ldots, G(A_r)) \sim \text{Dirichlet}(\alpha G_0(A_1), \ldots, \alpha G_0(A_r)) \tag{5.3.1}$$

for every finite measurable partition $A_1, \ldots, A_r$ over some probability space Θ.

The parameters G_0 and α play intuitive roles in the definition of the DP. The base distribution is basically the mean of the DP for any measurable set $A \subset \Theta$, that is,

$$\mathbb{E}[G(A)] = G_0(A).$$

On the other hand, the concentration parameter can be interpreted as an inverse variance

$$\mathbb{V}[G(A)] = \frac{G_0(A)(1 - G_0(A))}{\alpha + 1}.$$

The larger α is the smaller the variance and the DP will concentrate more of its mass around the mean. Now we are interested in the posterior distribution of G given some observed values. Let $\pi_1, \ldots, \pi_n$ be a sequence of independent draws from G. Note that the π_i's take values in Θ since G is a distribution over Θ. Let $A_1, \ldots, A_r$ be a finite measurable partition of Θ, and let n_k be the number of observed values in A_k. Then by the conjugancy between the Dirichlet and the multinomial distributions, we have

$$(G(A_1), \ldots, G(A_r)) \mid \pi_1, \ldots, \pi_n \sim \text{Dir}(\alpha G_0(A_1) + n_1, \ldots, \alpha G_0(A_r) + n_r). \tag{5.3.2}$$

Since the above is true for all finite measurable partitions, the posterior distribution over G must be a DP as well.

In fact, the posterior DP is

$$G \mid \pi_1, \ldots, \pi_n \sim DP\left(\alpha + n, \frac{\alpha G_0 + \sum_{i=1}^{n} \delta_{\pi_i}}{\alpha + n}\right).$$

Notice that the DP has updated concentration parameter $\alpha+n$ and base distribution $\frac{\alpha G_0 + \sum_{i=1}^{n} \delta_{\pi_i}}{\alpha+n}$, where δ_{π_i} is a point mass located at π_i and $n_k = \sum_{i=1}^{n} \delta_{\pi_i}(A_k)$. In other words, the DP provides a conjugate family of priors over distributions that are closed under posterior updates given observations.

Furthermore, notice that the posterior base distribution is weighted average between the prior base G_0 and the empirical distribution $\frac{\sum_{i=1}^{n} \delta_{\pi_i}}{n}$. Indeed, the weight associated with the prior base distribution is proportional to α, while the empirical distribution has weight proportional to the number of observations n.

Following we show how the posterior base distribution given $\pi_1, \pi_2, \dots \pi_n$ is also the predictive distribution of π_{n+1}. This is because of the conditional distribution of the DP we would expect that the probability of $\pi_1 = \pi_2$ is equal to $1/(\alpha+1)$. In fact the following results:

$$\begin{aligned}
\pi_1 &\sim G_0, \\
\pi_2 \mid \pi_1 &\sim \frac{\alpha}{\alpha+1} G_0 + \frac{1}{\alpha+1}\delta_{\pi_1}, \\
&\vdots \\
\pi_{n+1} \mid \pi_1, \dots, \pi_n &\sim \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n}\sum_{i=1}^{n}\delta_{\pi_i}
\end{aligned}$$

and the joint distribution of $\pi = \{\pi_1, \dots, \pi_n\}$ is

$$f(\pi) = \prod_{i=1}^{n} \frac{\alpha G_0 + \sum_{i=1}^{n} \delta_{\pi_i}}{\alpha + i - 1}.$$

Following we formalize these results as the Polya urn scheme of the DP prior (see Blackwell and MacQueen (1973)).

Definition 5.3.2. (Blackwell-MacQueen Urn Scheme) Let $G \sim DP(\alpha, G_0)$, and drawing an iid sequence $\pi_1, \pi_2, \dots \pi_n$. Consider the predictive distribution for π_{n+1} conditioned on $\pi_1, \pi_2, \dots \pi_n$ and with G marginalized out. Since

$$\pi_{n+1} \mid G, \pi_1, \pi_2, \dots \pi_n \sim G,$$

for a mesurable set $A \subset \Theta$ we have

$$\mathbb{P}(\pi_{n+1} \in A \mid \pi_1, \pi_2, \dots \pi_n) = \frac{1}{\alpha+n}\left(\alpha G_0(A) + \sum_{i=1}^{n} \delta_{\pi_i}(A)\right)$$

or with G marginalized out

$$\pi_{n+1} \mid \pi_1, \pi_2, \dots, \pi_n \sim \frac{1}{\alpha+n}\left(\alpha G_0(A) + \sum_{i=1}^{n} \delta_{\pi_i}(A)\right). \tag{5.3.3}$$

The name of urn model stems from the equivalence between the Dirichlet distribution and the Polya Urn scheme. Consider in particular an urn that at the outset contains a ball of a single color. At each step we either draw a ball from the urn and replace it with two balls of the same colour, or we are given a ball of a new colour which we place in the urn. The parameter α defines the probabilities of these two cases. Viewing each (distinct) color as a sample from G_0 and each ball as a sample from G, Blackwell and MacQueen (1973) showed that this Polya urn model yields samples whose distributions are those of the marginal probabilities under the Dirichlet process. The urn scheme also directly suggests a sampling based computational scheme for posterior inference.

In this chapter we are interested in modelling the density from which a given set of observations, the signatures patterns, is drawn. However, note that distributions drawn from a DP are discrete, thus do not have densities. Then a mixture of DP is the alternative.

There are several points worth noticing in DP mixture modeling via Pòlya urn schemes.

- First, as in Bayesian models for mixtures, the Pòlya urn scheme assumes exchangeability of the variables π_i.
- Second, scheme (5.3.3) shows that the DP exhibits a clustering property as a result of the discreteness property of the random measure G. Successive draws π_i from G have positive probability of being equal to one of the previous draws, and there is a positive reinforcement effect affecting future draws. Thus, scheme (5.3.3) reduces to a finite mixture model when the mass of G is concentrated on a few atoms.
- Third, scheme (5.3.3) emphasizes approximate predictive inference, and opens up the way to efficient MCMC sampling approaches for Bayesian density estimation based on posterior predictive distributions. Following this approach, Escobar and West (1995) were the first to propose efficient MCMC simulation methods for DP mixture models.
- Fourth, the DP mixture model produces predictive distributions qualitatively similar to traditional kernel techniques, but taking into account differing degrees of smoothing across the sample space.
- Finally, the Bayesian nonparametric approach supports a formal methodology for the analysis of nonparametric density estimation problems.

5.4. Dirichlet process mixtures of infinite Gaussian distributions for M_4 processes

In this section we describe multivariate Dirichlet process mixtures of infinite Gaussian distributions and key features of their structure in the context of M_4 processes estimation. Let $\boldsymbol{X}_i = \{X_{i,j}, \ldots, X_{i,d}\}$ be a $d-$dimensional M_4 process with underterminated number of signature patterns $\mathcal{L}$, but known cluster size $\mathcal{K}$. Furthermore, denote δ_u an indicator function of exccedances in at least one of the d- variate marginales over a threshold $\boldsymbol{u}$, i.e, $\boldsymbol{X} \nleq \boldsymbol{u}$. This indicator function takes the value 1 if an exceedance is observed or 0 otherwise. Define $N_u = \sum \delta_u$. Each time that a multivariate local maximum $\boldsymbol{X}_{i^*}$ with cluster size $\mathcal{K}$ is found we reorder this process to have a *standardized process of signature patterns* as follows:

$$\begin{pmatrix} X_{i^*-k,1} & \cdots & X_{i^*-k,d} \\ \vdots & & \vdots \\ X_{i^*-1,1} & \cdots & X_{i^*-1,d} \\ X_{i^*,1} & \cdots & X_{i^*,d} \\ X_{i^*+1,1} & \cdots & X_{i^*+1,d} \\ \vdots & & \vdots \\ X_{i^*+k,1} & \cdots & X_{i^*+k,d} \end{pmatrix}_{\mathcal{K}\times d} \Rightarrow \begin{pmatrix} \vdots & & & & \vdots \\ \frac{X_{i^*-k,1}}{X^{\Sigma}_{i^*}} & \cdots & \frac{X_{i^*-k,d}}{X^{\Sigma}_{i^*}} & \frac{X_{i^*-k+1,1}}{X^{\Sigma}_{i^*}} \cdots & \frac{X_{i^*+k,d}}{X^{\Sigma}_{i^*}} \\ \vdots & & & & \vdots \end{pmatrix}_{\mathcal{L}\times P},$$

where $X^{\Sigma}_{i^*}$ is the multivariate local maximum, $\mathcal{L} = N_u/\mathcal{K}$ and $P = d \cdot \mathcal{K}$.

For simplicity let us rewrite $\frac{X_{i^*-k,1}}{X^{\Sigma}_{i^*}}$ as an associated P-dimensional process

$$\boldsymbol{y}_l = (y_{l,1}, \ldots, y_{l,p}) = \left(\frac{X_{i^*-k,1}}{X^{\Sigma}_{i^*}}, \ldots, \frac{X_{i^*+k,d}}{X^{\Sigma}_{i^*}}\right) \tag{5.4.1}$$

for $l = 1, \ldots, \mathcal{L}$ and $p = 1, \ldots, P$.

Suppose that we observed $\boldsymbol{y}_l$ drawn from an uncertain distribution to be estimated. Let $\boldsymbol{\pi}_l^{\prime}$s be independent and have multivariate normal distributions with a vector of means $\boldsymbol{\mu}'$s and inverse covariance $\Omega_l^{\prime}$s, where $\boldsymbol{\pi}_l = (\boldsymbol{\mu}_l, \Omega_l)$. From a Bayesian perspective, density estimation may be viewed as a problem of predicting a further draw $\boldsymbol{y}_{\mathcal{L}+1}$; hence any model must provide a mean of computing the predictive distribution $\left(\boldsymbol{y}_{\mathcal{L}+1} \mid \boldsymbol{y}\right)$ for the further draw conditionally on $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{\mathcal{L}}\}$.

Escobar and West (1995) follow Ferguson (1983) in using a Dirichlet process model to define a class of normal mixture models for univariate density estimation. Here we have a multivariate generalisation of this previous work, in which we assume the following hierarchical description:

$$\begin{aligned}
\boldsymbol{y}_l \mid \boldsymbol{\mu}_l, \Omega_l &\sim N_P\left(\boldsymbol{\mu}_l, \Omega_l\right), \quad l = 1, \ldots, \mathcal{L} \\
\boldsymbol{\pi}_l = (\boldsymbol{\mu}_l, \Omega_l) \mid G &\sim G \\
\boldsymbol{\pi}_l = (\boldsymbol{\mu}_l, \Omega_{sl}) &\sim G \\
G \mid \alpha, G_0 &\sim DP\left(\alpha G_0\right).
\end{aligned} \tag{5.4.2}$$

In words, the discrete distribution G has a Dirichlet process prior with parameter αG_0, given G, the parameters $\boldsymbol{\pi}_l$ are independently drawn from G; then, given $\boldsymbol{\pi}_l$, the signature patterns $\boldsymbol{y}_l$ follow a multivariate normal distribution with moments given by components of $\boldsymbol{\pi}_l$. An additional level may be added to this hierarchy framework to specify hyperpriors on the parameters α and G_0. Indeed, this model can be extended to a richer class by putting a hyperprior on α and in the hyperparameters of G_0, see Escobar and West (1995) for such extensions in univariate models.

We complete our model by stating distributions on the hyperparameters. Under G_0, $\boldsymbol{\mu}_l$ and Ω_l are independent; $\boldsymbol{\mu}_l$ has a multivariate normal prior and Ω_l a Wishart prior. Here $W_P\left(\cdot \mid v_1, \psi^{-1}\right)$ denotes a P-dimensional Wishart distribution with v_1 degrees of freedom and the $P \times P$ positive definite scale matrix ψ. In resume

$$\begin{aligned}
G_0 &= f\left(\boldsymbol{\mu}_l, \Omega \mid \boldsymbol{m_1}, \Sigma/\kappa, v_1, \psi^{-1}\right) & (5.4.3)\\
&= f\left(\boldsymbol{\mu}_l \mid \boldsymbol{m_1}, \Sigma/\kappa\right) f\left(\Omega \mid v_1, \psi^{-1}\right), \\
f\left(\boldsymbol{\mu}_l \mid \boldsymbol{m_1}, \Sigma/\kappa\right) &\sim N_P\left(\boldsymbol{\mu}_l \mid \boldsymbol{m_1}, \kappa/\Sigma\right) \\
\boldsymbol{\mu}_l &\sim N_P\left(\boldsymbol{m_1}, \kappa/\Sigma\right) & (5.4.4)\\
&= (2\pi\kappa)^{P/2} \left|\Sigma\right|^{1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_l - \boldsymbol{m_1}\right)^T \left(\Sigma/\kappa\right)\left(\boldsymbol{\mu}_l - \boldsymbol{m_1}\right)\right\}, \\
f\left(\Omega_l \mid v_1, \psi^{-1}\right) &\sim W_P\left(v_1, \psi^{-1}\right) \\
\Omega_l &\sim W_P\left(v_1, \psi^{-1}\right) & (5.4.5)\\
&= \frac{|\psi|^{v_1/2}}{2^{v_1 P/2}\Gamma_P\left(v_1/2\right)} \left|\Omega_l\right|^{(v_1-P-1)/2} \exp\left\{-\frac{v_1}{2}\mathrm{trace}\left(\Omega_l \psi\right)\right\},
\end{aligned}$$

with $v_1 > P - 1$, $l = 1, \ldots, L$, where $\boldsymbol{m_1}, \Sigma, \kappa, v_1$ and ψ are the hyperparameters and the Wishart prior is parametrized such that if $\Psi \sim W_P\left(\cdot \mid v_1, \psi^{-1}\right)$ then $E\left(\Psi\right) = \psi^{-1}/\left(v - P - 1\right)$.

Moreover, $\Gamma_P(x)$ is the P-variate generalised Gamma function

$$\Gamma_P(x) = \pi^{P(P-1)/4} \prod_{j=1}^{P} \Gamma\left(x + \frac{(j-P)}{2}\right), \quad x > (P-1)/2.$$

Each Ω_s, Σ, and ψ is a $P \times P$ positive-definite symmetric matrix. We assume the following priors for hyperparameters $\boldsymbol{m_1}, \Sigma$ and ψ assuming mutually independence;

$$\begin{aligned} \boldsymbol{m_1} \mid m_2, \Delta &\sim N(\boldsymbol{m_2}, \Delta) \\ \kappa \mid \tau_1, \tau_2 &\sim \Gamma(\tau_1/2, \tau_2/2) \\ \psi \mid v_2, \psi_2 &\sim W_P\left(v_2, \psi_2^{-1}\right), \end{aligned}$$

where $\Gamma(\tau_1, \tau_2)$ denotes a gamma distribution with shape τ_1 and scale τ_2. We also have the hyperparameter α as the concentration parameter of the underlying Dirichlet process specified as

$$\alpha \mid \varrho_0, \varrho_1 \sim \Gamma(\varrho_0, \varrho_1).$$

Under the described model, posterior computations via Markov chain simulation are feasible. The univariate development in Escobar and West (1995) and Rasmussen (2000) can be extended and modified to produce various methods of estimation described below.

5.5. Estimation of Dirichlet processes mixtures

The use of DPM models has become computationally feasible with the development of Markov chain methods for sampling from the posterior distribution of the parameters of the component distributions and of the associations of mixture components with observations. Methods based on Gibbs sampling can easily be implemented for models based on conjugate prior distributions, but when nonconjugate priors are used, as is appropriate in many contexts, straightforward Gibbs sampling requires that an often difficult numerical integration be performed.

Neal (2000) presented several possible algorithms for sampling from the posterior distribution of Dirichlet process mixtures. In this research, we use Gibbs sampling with auxiliary parameters (Neal's algorithm 8). This approach is similar to the algorithm proposed by MacEachern and Mueller (1998), with a difference that the auxiliary parameters exist only temporarily.

We begin demonstrating that the model (5.3.3) can also be obtained by taking the limit as $l' \to L$ goes to ifinity of mixture of normal models with l' components having the following form

$$f(\boldsymbol{y}) = \sum_{c=1}^{l'} p_c f(\boldsymbol{y} \mid \phi_c). \tag{5.5.1}$$

Here, p_c are the *mixing proportions* (which must be positive and sum to one), and f is a simple class of distributions, in our case Gaussian as in (5.4.2) with $\phi_{c_l} = \boldsymbol{\pi}_l = (\boldsymbol{\mu}_l, \Omega)$.

Given the set $\boldsymbol{y}$ the classical approach to estimate the parameters $(\boldsymbol{\mu}_l, \Omega_l)$, is to maximize the likelihood by using the EM algorithm (Dempster et al. (1977)). The EM algorithm guarantees convergence to a local maximum, with the quality of the maximum being heavily dependent on the random initialization of the algorithm. Alternatively, a Bayesian approach can be used to combine the prior distribution for the parameters and the likelihood, resulting in a joint posterior distribution:

$$f(\mu, \Omega, p \mid \boldsymbol{y}) \propto f(\mu, \Omega, p)\, f(\boldsymbol{y} \mid \mu, \Omega, p)\,. \tag{5.5.2}$$

However, the joint posterior takes a highly complicated form. Thus, it is generally not feasible to perform analytical inference based on the above posterior distribution. The MCMC approaches have typically been used to calculate the joint posterior and, of the approaches that have been proposed in the literature, Gibbs sampling is suitable for mixture models.

We first assume that the number of mixing components, l', is finite, and later we will use the model in the limit where $l' \to \mathcal{L} = \infty$. As a prior for p_l is a symmetric Dirichlet distribution:

$$\mathbb{P}(p_1, \ldots, p_{l'}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^{l'}} \prod_{c=1}^{l'} p_c^{(\alpha/l')-1}$$

where $p_c \geq 0$ and $\sum p_c = 1$. Parameters ϕ_c are assumed to be independent under the prior with distribution G_0, while the value α is assumed at the moment known. We can use *mixture identifiers*, c_l, which are stochastic variables whose values encode the class to which observation $\boldsymbol{y}_l$ belongs. The actual values of the identifiers are arbitrary, as long as they faithfully represent which observations belong to the same classes, but can be thought of as taking values from $1, \ldots, l'$, where l' is the total number of classes observed. Neal (2000) represents the above mixture model as follows

$$\begin{aligned}
\boldsymbol{y}_l \mid \boldsymbol{c}, \phi &\sim F(\phi_{c_l}) \\
\boldsymbol{c_l} \mid p &\sim \text{Discrete}(p_1, \ldots, p_{l'}) \\
\phi_{\boldsymbol{c}} &\sim G_0 \\
p_1, \ldots, p_{l'} &\sim \text{Dirichlet}\left(\alpha/l', \ldots, \alpha/l'\right).
\end{aligned} \tag{5.5.3}$$

Now we define the *occupation numbers* $n_j = \sum_{l=1}^{\mathcal{L}} \delta(c_l, j)$ as the number of observations associated with the component j. This definition allow to define the prior (multinomial) for the occupation numbers given the mixing proportions

$$\begin{aligned}
p(n_1, \ldots, n_j \mid p_1, \ldots, p_{l'}) &= \frac{\mathcal{L}!}{\prod_{j=1}^{l'} n_j!} \prod_{j=1}^{\mathcal{L}} p_j^{n_j} \\
p(c_1, \ldots, c_{\mathcal{L}} \mid p_1, \ldots, p_{l'}) &= \prod_{j=1}^{C} p_j^{n_j}
\end{aligned}$$

By integrating over the Dirichlet prior, we can eliminate mixing proportions, p_c. This is an important step for the estimation of this kind of processes. First, notice that we can write the prior directly in terms of indicators:

$$\begin{aligned} f(c_1,\dots,c_{\mathcal{L}} \mid \alpha) &= \int f(c_1,\dots,c_L \mid p_1,\dots,p_{l'}) f(p_1,\dots,p_{l'})\, dp_1\cdots dp_{\mathcal{L}} \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha/l')^{l'}} \int \prod_{j=1}^{l'} p_j^{n_j+\alpha/l'-1} \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha+\mathcal{L})} \prod_{j=1}^{l'} \frac{\Gamma(n_j+\alpha/l')}{\Gamma(\alpha/l')}. \end{aligned} \tag{5.5.4}$$

Fixing all but a single indicator c_l in (5.5.4) we obtain the following conditional distribution for c_l:

$$\mathbb{P}(c_l = c \mid c_1,\dots,c_{l-1},\alpha) = \frac{n_{-lc}+\alpha/l'}{i-1+\alpha}. \tag{5.5.5}$$

Here, n_{-lc} represents the number of data points previously (i.e., before the l-th pattern) assigned to component c, excluding $\boldsymbol{y}_l$. The probability of assigning each component to the first data point is $1/l'$. As we proceed, this probability becomes higher for components with larger numbers of samples (i.e., larger $n_{-l'c}$). When $\mathcal{L}$ goes to infinity, the conditional probabilities (5.5.5) reach the following limits:

$$\mathbb{P}(c_l = c \mid c_1,\dots,c_{l-1}) \to \frac{n_{-lc}}{l-1+\alpha} \tag{5.5.6}$$

$$\mathbb{P}(c_l \neq c_j\ \forall j < l \mid c_1,\dots,c_{l-1}) \to \frac{\alpha}{l-1+\alpha}. \tag{5.5.7}$$

As a main result, the conditional probability for $\pi_l = \phi_{c_l}$, becomes relation (5.3.3)

$$\pi_i \mid \pi_1,\pi_2,\dots,\pi_{l-1} \sim \frac{1}{\alpha+l-1}\left(\alpha G_0 + \sum_{j<l}\delta_{\pi_j}\right), \tag{5.5.8}$$

which is equivalent to DP mixture model (5.4.2).

Note as the results in (5.5.6) and (5.5.7) show that the conditional class prior for classes that are associated with observations other than $\boldsymbol{y}_l$, is proportional to the number of such observations and that the combined prior for all other classes depends only on α and l'.

The main advantage of this representation is that it allows us to work with the finite number of indicators c_l, rather than the infinite number of mixing proportions, which is crucial for the estimation of this class of processes.

5.5.1. Gibbs sampling using a DPM. The simplest case arises when a conjugate prior is used. In the terminology of the DP, this means that the data sampling distribution G is conjugate to the base density G_0 of the DP. To perform inference with conjugate priors, we need to be able to compute the marginal distribution of a single observation and need to be able to draw samples from the posterior of the base distributions. For instance, the most direct approach to sampling for model (5.5.3) is to repeatedly draw values for each π_l from its conditional distribution given both the data $\boldsymbol{y}_l$ and the $\pi_{-l} = \{\pi_1,\dots,\pi_{l-1},\pi_{l+1},\dots,\pi_{\mathcal{L}}\}$. This conditional distribution is obtained by combining the likelihood, i.e. $f(\boldsymbol{y}_l \mid \pi_l)$, and the prior conditional on π_{-l} which is

$$\pi_l \mid \pi_{-l} \sim \frac{1}{L-1+\alpha} \sum_{j \neq l} \delta_{\pi_j} + \frac{\alpha}{L-1+\alpha} G_0,$$

since the observations are exchangeable. Combined with the likelihood (5.5.1), this yields the following conditional distribution for use in Gibbs sampling:

$$\pi_l \mid \pi_{-l}, \boldsymbol{y}_l \sim \sum_{j \neq l} q_j \delta_{\pi_j} + r_l G_l,$$

where G_l is the posterior distribution for π based on the prior G_0 and the observation $\boldsymbol{y}_l$, with likelihood $f(\boldsymbol{y}_l \mid \pi)$. The values of the q_j and of r_l are defined as

$$\begin{aligned} q_j &= b \quad f(\boldsymbol{y}_l \mid \pi_j) \\ r_l &= b\alpha \int f(\boldsymbol{y}_l \mid \pi)\, dG_0(\pi) \end{aligned}$$

where b_l such that $\sum_{j \neq l} q_j + r_l = 1$ and

$$q_j \propto N\left(\boldsymbol{y}_l \mid \boldsymbol{\mu}_j, \Omega_j\right) = |\Omega_j|^{1/2} \exp\left\{-\Omega_j \left(\boldsymbol{y}_l - \boldsymbol{\mu}_l\right)^2 / 2\right\}.$$

For this Gibbs sampling method to be feasible, computing the integral defining r_l and sampling from G_l must be feasible operations.

We can improve the estimation if the mixing proportions p_c are integrated out in (5.5.3) and combined with the likelihood given by $f(\boldsymbol{y}_l \mid \pi)$, but first we need to update each component of G_0 conditional to the observations $\boldsymbol{y}_l$.

5.5.2. Conditional posterior distribution of the mixture parameters. Since the conditional posterior are of standard form for $\boldsymbol{\mu}_j$ and Ω_j, these can easily be updated using Gibbs sampling. We have to combine simply equations (5.4.2) and (5.4.3). As consequence of this framework, the conditional posterior for the parameters are also Gaussian and Wishart, respectively

$$\begin{aligned} f\left(\boldsymbol{\mu}_j \mid \mathbf{c}, \boldsymbol{y}, \Omega_j, \boldsymbol{m_1}, \Sigma\right) &\sim N_P\left(\frac{n_j \Omega_j \bar{\boldsymbol{y}}_j + \Sigma \boldsymbol{m_1}}{n_j \Omega_j + \Sigma}, \frac{1}{n_j \Omega_j + \Sigma}\right), \\ f\left(\Omega_j \mid \mathbf{c}, \boldsymbol{y}, v_1, \psi\right) &\propto |\Omega_j|^{(v_1 - P - 1)/2} \exp\left\{-\frac{v_1}{2} \operatorname{trace}\left(\Omega_j \psi\right)\right\} \\ & \quad \prod_{i: c_i = j} |\Omega_j|^{1/2} \exp\left(-\frac{1}{2}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_j\right)^T \Omega_j \left(\boldsymbol{y}_i - \boldsymbol{\mu}_j\right)\right) \\ &\sim W_P\left(v_1 + n_j, \frac{1}{v_1 \psi + n_j S_j\left(\boldsymbol{y} \mid \boldsymbol{\mu}_j\right)}\right), \end{aligned}$$

where $\bar{\boldsymbol{y}}_j = n_j^{-1} \sum_{l: c_l = j} \boldsymbol{y}_l$ is the mean and $S_j\left(\boldsymbol{y} \mid \boldsymbol{\mu}_j\right) = n_j^{-1} \sum_{l: c_l = j} \left(\boldsymbol{y}_l - \boldsymbol{\mu}_j\right)\left(\boldsymbol{y}_l - \boldsymbol{\mu}_j\right)^T$ is the sample covariance of the observations belonging to component j given the mean $\boldsymbol{\mu}_j$.

5.5.3. Conditional posterior distribution of the mixing weights. In the representation (5.5.3) we observe that the main advantage is that this allows us to work with a finite number of indicators, rather than the infinite number of mixing proportions. However, when $l^{'} \rightarrow \mathcal{L} = \infty$ we cannot explicitly represent the infinite number of ϕ_c. We instead represent (and do Gibbs sampling for) only those ϕ_c that are currently associated with some observation. Gibbs sampling for the c_l is based on the following conditional probabilities

$$\begin{aligned}
\mathbb{P}\left(c_l = c \mid c_{-l}, \boldsymbol{y}_i, \phi\right) &= b\frac{n_{-i,c}}{L-1+\alpha} \quad f\left(\boldsymbol{y}_i \mid \phi_c\right) \text{ if } c = c_j \\
&\quad \text{for some } j \neq i.
\end{aligned} \tag{5.5.9}$$

$$\mathbb{P}\left(c_l = c \,\forall j \neq i \mid c_{-i}, \boldsymbol{y}_i, \phi\right) = b\frac{\alpha}{L-1+\alpha}\int f\left(\boldsymbol{y}_i \mid \phi\right) dG_0\left(\phi\right), \tag{5.5.10}$$

where ϕ is the set of ϕ_c currently associated with at least one observation and b is the appropriate normalizing constant that makes the above probabilities sum to one. Furthermore in our approach

$$\begin{aligned}
f\left(\boldsymbol{y}_l \mid \phi_c\right) &\propto N_P\left(\boldsymbol{y}_l \mid \boldsymbol{\mu}_j, \Omega_j\right) = \left|\Omega_j\right|^{1/2} \exp\left\{-\Omega_j\left(\boldsymbol{y}_l - \boldsymbol{\mu}_l\right)^2/2\right\} \\
\int f\left(\boldsymbol{y}_l \mid \phi\right) dG_0\left(\phi\right) &\propto \int f\left(\boldsymbol{y}_l \mid \phi\right) f\left(\boldsymbol{\mu}_j, \Omega \mid \boldsymbol{m_1}, \Sigma/\kappa, v_1, \psi^{-1}\right).
\end{aligned}$$

This is essentially the method used by Escobar and West (1995). The approach is feasible when we can compute the integral $\int f\left(\boldsymbol{y}_l, \phi\right) dG_0\left(\phi\right)$ or use a conjugate prior.

5.5.4. Infinite Gaussian mixture model. The previous discussions have been restricted to a finite number of mixtures. From now on we adopt the perspective from the DPM, assuming the number of mixtures tends to be ∞.

Notice the particular consequence of this idea on the estimation of M_4 processes. We mentioned in section (5.2) that we can approximate a M_4 process by other finite M_4 process. Even there we are not sure that we can observe all signature patterns from the real process, but we hope to obtain enough information as soon as the simple size increases infinitely. Indeed, the DPMs assume the existence of infinite patterns in our case, therefore, there must be an infinite number of mixtures with no patterns $\boldsymbol{y}$ associated with them. These are termed "unrepresented" mixtures. Correspondingly, "represented mixtures" are those that have training patterns $\boldsymbol{y}$ that are associated with them.

Let k^- denote the number of represented mixtures. For represented mixtures, the previously derived conditional posteriors of (μ_l, Ω_l) in (5.4.4) and (5.4.5) still hold. In contrast, in the absence of training data, the parameters in unrepresented mixtures are solely determined by their priors $f\left(\boldsymbol{\mu}_l \mid \boldsymbol{m_1}, \Sigma/\kappa\right)$ and $f\left(\Omega \mid v_1, \psi^{-1}\right)$. Thus the inference of the indicators, c_l, must incorporate the effect of infinite mixtures. This is done as in equations (5.5.6) and (5.5.7) for the represented and not represented mixtures respectively.

The conditional posteriors of (μ_l, Ω_l) are Gaussian and Wishart distributions respectively, from which samples can be generated by using standard procedures. The sampling of the indicators requires the evaluation of the integral in equation (5.5.7), which is only analytically feasible if the conjugate prior is used. To approximate this integral we use the algorithm 8 by Neal (2000) which essentially samples just a few auxiliary variables in a way that does not affect the detailed balance condition that guarantees that the outer Markov chain converges to the correct stationary distribution.

The overall structure of the sampling algorithm remains identical in the case of non-conjugate priors. However, the sampling for the indicator variables c_l changes slightly when c_l is updated. In this case we will introduce m temporary auxiliary variables that represent possible values for the parameters of components that are not associated with any other observations. We then update c_l by Gibbs sampling with respect to the distribution that includes these auxiliary parameters.

Because of the fact that the observations $\boldsymbol{y}_l$ are exchangeable, and the component labels c_l are arbitrary, we can assume that we are updating c_l for the last observation, and that the c_j for other observations have values in the set $\{1, \ldots, k^-\}$, We can now visualize the conditional prior distribution for c_l given the other c_j in terms of these m auxiliary components and their associated parameters. The probability of c_l being equal to a c in $\{1, \ldots, k^-\}$ will be $n_{-l,c}/\left(\mathcal{L} - 1 + \alpha\right)$, where $n_{-l,c}$ is the number of times c occurs among the c_j for $j \neq l$. The probability of c_l having some other value will be $\alpha/\left(\mathcal{L} - 1 + \alpha\right)$ which we will split equally among the m auxiliary components we have introduced.

The first step in using this representation to update c_l is to sample from the conditional distribution of these auxiliary parameters given the current value of c_l and the rest of the state. In the case of the conditional probability (5.5.9) (i.e., $n_{-l,c} > 0$) the auxiliary parameters have no connection with the rest of the state or the observations, and are simply drawn independently from G_0. In the case the conditional probability (5.5.10) (i.e., $n_{-l,c} = 0$), observation $\boldsymbol{y}_l$ is currently the only observation associated with the class j, which means it must be associated with one of the m auxiliary parameters. We select the first auxiliary parameter c_l with the corresponding value ϕ being equal to the existing ϕ_{c_l}. The values for the other auxiliary components are drawn independently from G_0 again. In our investigation we set $m = 1$, therefore, this step will not be necessary.

Following, a Gibbs sampling update is performed for c_l in this representation of the posterior distribution. Since c_i must be either one of the components associated with other observations or one of the auxiliary components that were introduced, we can easily do Gibbs sampling by evaluating the relative probabilities of these possibilities. Once a new value for c_l has been chosen, we discard all values that are not associated with an observation.

Notice that all classes, existing as well as auxiliary, have parameters associated with them. Thus, we can evalute their likelihoods and the priors without problems, which takes the form $\frac{n_{-l,c}}{\mathcal{L}-1+\alpha}$ for components with observations other than $\boldsymbol{y}_l$ associated with them and $\frac{\alpha/m}{\mathcal{L}-1+\alpha}$ for auxiliary classes. We summarize Algorithm 8:

ALGORITHM 5.5.1. *(Algorithm 8 in Neal (2000)). Let the state of the Markov chain consist of* $c_1, \ldots, c_{\mathcal{L}}$ *and* $\phi = \{\phi_c : c \in (c_1, \ldots, c_{\mathcal{L}})\}$. *Repeatedly sample as follows:*
for *(i in* $1 : n$*)*

(1) ***Set***
    k^- *= the number of distinct* c_j *for* $j \neq l$,
    $h = k^- + m$,
    $n_{-l,c}$ *= the number of* c_j *for* $j \neq l$ *that are equal to* c,
    b *is the appropriate normalizing constant.*
(2) ***Label*** c_j *with values in* $\{1, \ldots, k^-\}$
    (a) *If* ($c_l = c_j$) *for some* $j \neq l$
        (i) *draw values independently from* G_0 *for* c_j *for which* $k^- < c \leq h$.
    (b) *If* ($c_l \neq c_j$) *for all* $j \neq l$,

(i) *set c_l the label $k^- + 1$, and*

(ii) *draw values independently from G_0 for those ϕ_c for which $k^- + 1 < c \leq h$.*

(3) ***Draw*** *a new value for c_l from $\{1, \dots, h\}$ using the following probabilities:*

$$\mathbb{P}(c_l = c \mid c_{-l}, \boldsymbol{y}_l, \phi_1, \dots, \phi_{\mathcal{L}}) = \begin{cases} b\frac{n_{-il,c}}{\mathcal{L}-1+\alpha} f(\boldsymbol{y}_l \mid \phi_c) & \text{for } 1 < c \leq k^- \\ b\frac{\alpha/m}{\mathcal{L}-1+\alpha} f(\boldsymbol{y}_l \mid \phi_c) & \text{for } k^- < c \leq h \end{cases}$$

(a) *Change the state to contain only those ϕ_c that are now associated with one or more observations.*

(4) ***For all*** *$c \in (c_1, \dots, c_{\mathcal{L}})$:*

(a) *Draw a new value from $\phi_c \mid \boldsymbol{y}_l$ subject to $c_l = c$, or*

(b) *Perform some other update to ϕ_c that leaves this distribution invariant.*

5.5.5. Prediction. The calculation of the predictive probability of new data will be averaged over a number of MCMC samples, which are selected from those samples where the algorithm tends to stabilize. Stabilization will be assessed heuristically based on the value of the log-likelihood. Additionally to eliminate the auto-correlation, one sample will be selected from each consecutive set of 10 iterations. For a particular MCMC sample, the predictive probability is attained from two components: the represented and the unrepresented mixtures. In a similar manner to that adopted in the sampling stage, the probability from unrepresented mixtures will be approximated by a finite mixture of Gaussians, whose parameters $(\boldsymbol{\mu}_l, \Omega_l)$ are drawn from the prior.

5.6. Practical implementation and simulation examples

In this section, we adress the practical implementation and the accuracy of the proposed model. The case of unknown cluster size coefficient $\mathcal{K}$ is addressed in Section 5.7 with some practical applications.

In Section 5.2 we have introduced the concept of local maximum and signature patterns, which allow estimating the models. Of course in real applications we do not expect the data to follow exactly a M_4 process, because the degenerate features of the repeated signature patterns are not likely to be a real case. Nevertheless, we hope that with large enough $\mathcal{L}$ and $\mathcal{K}$, the model should provide a good approximation of a finite dimensional max-stable process and therefore, to a wide class of multivariate time series.

Consider the process proposed in (5.2.6). In Section 5.2 we have shown for M_4 processes that the relative frequency of the $\mathcal{L}$-th signature pattern n_l is proportional to $\sum_{|k|} \max_d a_{lkd}$. This is the key feature that is used for identifying the parameters in a M_4 process by means of the standardized signature pattern given by equation (5.4.1) and explained extensively in section 5.4.

Before we give a series of experiments to investigate the quality of our approach it is helpful to explain at least one example in more detail. Let us consider the simulation of a M_4 process in 5-dimensions with 7 signature patterns and $\mathcal{K} = 2$. This means that the real dimension of the data will be 10-dimensions. The coefficients a_{lkd} were generated randomly from an uniform distribution. The sample size is 50.000 and the sample over a threshold $\boldsymbol{u}$ equals the 0.95-quantile. Convergence of the chain is determined by visual inspection of the trace

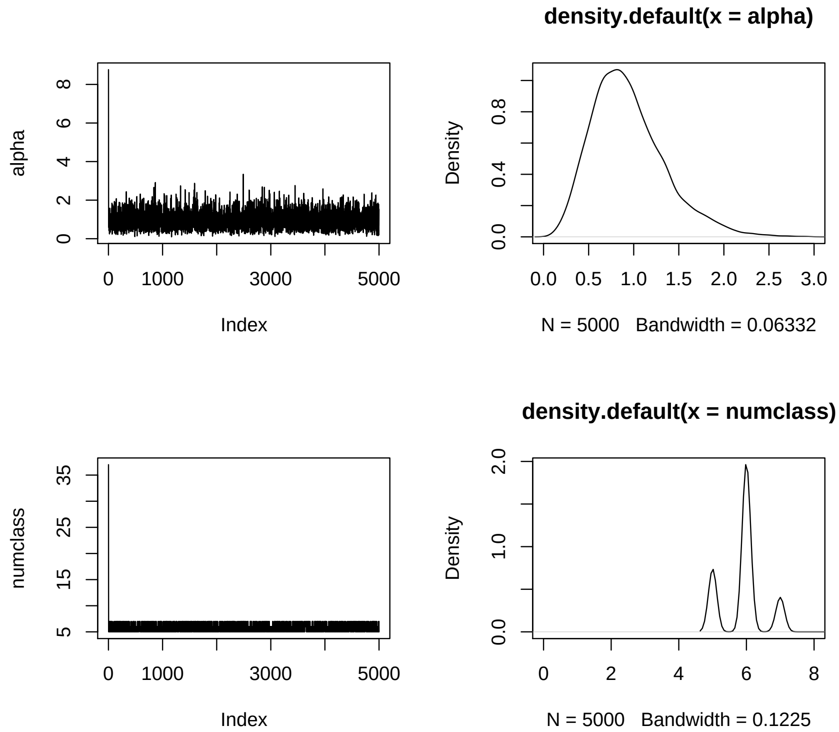

FIGURE 5.6.1. Monte Carlo Markov chain iterations for the patameter α (top panel) and the number of represented mixtures (bottom panel) together with the empirical density estimated for the results.

plots. However, based on the outputs of the chains, it is possible to use for instance the coda (Plummer et al. (2006)) R packages to perform convergence diagnostics. For instance a total of 5.000 iterations are investigated of which the first 1.000 are discarded. The remaining 4.000 are grouped in batches of 10 each, with the batch means providing the posterior estimates for the cluster variables.

In Figures 5.6.2, 5.6.3 and 5.6.4 the results are displayed graphically. The $\times$ in black colours are the true signature patterns $\boldsymbol{y}_p = \{y_{1p}, \dots y_{l'p}\}$, while the points in grey colours are the empirical estimations of them, which are influenced by the Multivariate Gaussian noise by assumption. The axes represent the number of possible combinations between all P-dimensions. In this case $\mathcal{K}{\cdot}d = 10$, which gives a total of 45 bidimensional faces. The ellipsoids, i.e. the linear transformations of the hyperspheres of the densities of the multivariate normal distributions, displayed in the figures are given by the eigenvectors of the covariance matrix Ω. The squared relative lengths of the principal axes are given by the corresponding eigenvalues. Thus, we can observe graphically how the Dirichlet mixture tries to fit the data.

The result of this fit is very encouraging since it shows that our model leads to a density profile that matches quite closely those of observed clusters. There are some more clusters than the real number of signature patterns, but it is natural in relation to the difficulty to identify the signature patterns plus a noise process. Furthermore, these extra clusters only have few observations, so that their influence in the determination of the coefficients a_{lkd} is minimal.

The top panel in Figure 5.6.1 illustrates the Monte Carlo Markov chain (MCMC) iterations of the parameter α of the DP mixture and the density of the estimates. The bottom panel shows the number of represented mixtures in relation to the number of MCMC iterations. The results of the simulation show that 7 mixtures were automatically inferred from the data. The parameter α estimate was 1.227 with 95% confidence bounds 1.720 and 0.378 given by the empirical quantile of the MCMC iterations.

Other important factors which naturally influence the results are the sample size, the number of observations over the threshold $\boldsymbol{u}$ and the distance among the signature clusters. For this example the minimal, maximal and mean Euclidean distance among the centre of the patterns are 0.548, 0.923 and 0.743 respectively. A more precise statistic to measure the distance among Gaussian distributions in a multivariate framework is the c-separation. This measure gives us more detailed information based not only on distance among the mean of the mixtures, but on the structure of the distributions.

Two Gaussian distributions $N_1(\mu_1, \Omega_1)$ and $N_2(\mu_2, \Omega_2)$ in a P-dimensional space are c-separated if

$$\|\mu_1 - \mu_2\| \geq c\sqrt{n \max\left(\lambda_{max}(\Omega_1), \lambda_{max}(\Omega_2)\right)},$$

where $\lambda_{max}(\Omega)$ is shorthand for the largest eigenvalue of Ω. A mixture of Gaussians is c-separated if its component Gaussians are in pairwise c-separated. Values lower than 1 means that the mixture of Gaussians overlaps significantly, while higher values correpond to almost completely separated Gaussian distributions. Thus, this definition gives us a theoretical framework to characterize the distance among cluster and therefore the difficulty to estimate the signature patterns. In the example the c-separation among the mixture of Gaussian distributions estimated is 0.018, which means that at least two multivariate Gaussian distributions are very close. Indeed, Figures 5.6.2, 5.6.3 and 5.6.4 illustrate that there are two clusters which are not easy to distinguish between them.

Now we realize some experiments of reduced sample size to investigate the influence of the number of patterns $\mathcal{L}$ in the sample in comparison to the number of dimensions. In the first experiment we simulated a sample of size 10.000 of a M_4 process with fixed cluster size $\mathcal{K} = 2$, and dimension $d = 3$ plus a d-dimensional Gaussian noise with mean 0 and covariance matrix

$$cov = \begin{pmatrix} 1 & 0.05 & \cdots & 0.05 \\ 0.05 & \ddots & & \vdots \\ \vdots & & \ddots & 0.05 \\ 0.05 & \cdots & 0.05 & 1 \end{pmatrix}_{d \times d}.$$

The number of signature pattern are $\mathcal{L} = 2, 3, 5, 10, 15$.

In the second experiment we kept the size of the sample, the cluster size $\mathcal{K} = 2$ but now we fixed the number of signature patterns to $\mathcal{L} = 5$. The coefficients a_{lkd} are simulated from a uniform distribution between 0 and 1 and they are standardized so that the sum of all is 1 for each dimension. The most optimistic scenario for estimating l' is one in which the observed points fall into distinct clusters sufficiently well separated (larger c-separation) in P- dimensions.

Scenario for dimension $d = 3$								
$\mathcal{L}$	N_u	l'	α	MSE	MIN	MAX	MEAN	c-separation
3	176	5	0.374	0.0121	0.238	0.840	0.631	0.0338
5	287	6	1.995	0.0192	0.368	1.033	0.727	0.0111
7	303	7	1.413	0.0323	0.170	0.865	0.593	0.0318
10	237	6	1.306	-	0.357	1.007	0.669	0.0452
15	258	9	1.815	-	0.209	0.967	0.617	0.0434

TABLE 5.6.1. Results for the first experiment, varying the number of true patterns $\boldsymbol{y}_l$, and keeping d and $\mathcal{K}$ fixed at 3 and 2 respectively. $\mathcal{L}$ is the true number of signature patterns. N_u the number of exceedances. l' the estimated or present number of signature patterns. MSE, MIN, MAX and MEAN are distance statistics based on the true coefficients a_{lkd}. The c-separation is calculated from the estimated signature patterns.

Scenario for patterns $\mathcal{L} = 5$							
d	l'	α	MSE	MIN	MAX	MEAN	c-separation
3	6	1.9956	0.0231	0.368	1.033	0.727	0.0111
5	5	0.7016	0.0245	0.498	0.952	0.757	0.0430
7	6	0.623	0.0301	0.555	1.032	0.858	0.0159
10	5	1.0657	0.0331	0.691	0.978	0.859	0.0058
15	5	0.9681	0.0392	1.015	1.319	1.189	0.0069

TABLE 5.6.2. Results for the second experiment, varying the number of dimensions d, and keeping $\mathcal{L}$ and $\mathcal{K}$ fixed at 5 and 2 respectively. $\mathcal{L}$ is the true number of signature patterns. N_u the number of exceedances. l' the estimated or present number of signature patterns. MSE, MIN, MAX and MEAN are distance statistics based on the true coefficients a_{lkd}. The c-separation is calculated from the estimated signature patterns.

To get some insights of the results, we studied the accuracy of each scenario as a function of the distance among the true standardized patterns $\boldsymbol{y}_l$ and the estimated ones. We calculate the minimal, maxima and mean distance among them in relation to the Euclidean distance. In Tables 5.6.1 and 5.6.2 we show the results of the two experiments. In Table 5.6.1 we vary the number of true patterns $\boldsymbol{y}_l$ (keeping $d = 3$ and $\mathcal{K}$ fixed at 2) while in 5.6.2 we despicte the results as we vary the number of dimensions d (keeping $\mathcal{L}$ fixed at 5). The error among the estimation of the patterns $\boldsymbol{y}_l$ and the true value are computed as the Mean squared error (MSE). Further, we calculate the c-separation for the density of mixture estimated. In all experiments we work with the 0.95-quantile as the threshold level $\boldsymbol{u}$. In the implementation we need to determine the clusters among the dimensions, so that the clusters do not get the values 1 or 0, which will produce clusters on the axes, we take all maximum locals, where one exceedance is observed in at least two marginals X_j.

In the first experiment the results are mixed. For a reduced number of signatures, the approach gives an accuracy estimation, while when the number of signatures is in the order of $\mathcal{L} \geq 10$, there are big difference in the estimations. The reason is that we cannot hope to observe all signature patterns $\mathcal{L}$ in a reduced number of standardized data $\boldsymbol{y}_l$, which are not enough to get a good estimation of the M_4 coefficients. Although, as we illustrate in the above example, we can get good estimates of the M_4 parameters when the sample is larger. Thus, the accuracy of the model depends considerably on the size of the sample.

In the second experiment the results are more homogeneous. We observe in general that although the number of dimensions caused a difficulty to estimate the correct number of signature patterns, this is not as critical as the number of patterns $\mathcal{L}$. Notice the c-separation is reduced as the number of dimensions increases, but the approach still giving a good estimation.

From the results in the experimets we can conclude that DP mixtures are able to give an accurate estimation of the signature patterns in a M_4 process. We observe further, that the estimation of the exact number of patterns $\mathcal{L}$ depends more on the number of exceedances over a threshold $\boldsymbol{u}$, while the number of dimensions plays a minor role.

An aspect that we do not study here, is the influence of the cluster size $\mathcal{K}$. On the one hand, in the proposed framework this factor influences only the number of dimensions P. From the point of view of estimation, a model with cluster size 2 and 3-dimensional is equivalent to a model with cluster size 3 and 2-dimensional. On the other hand, if the cluster size $\mathcal{K}$ is larger, then the number of standarized signature patterns $\boldsymbol{y}_l$ should be lower, by the fact that the dimension of $\boldsymbol{y}_l$ is $\frac{N_u}{\mathcal{K}} \times P$.

5.7. Applications in Finance

The subprime mortgage financial crisis is an ongoing crisis which was caused by the sharp rise in the US subprime mortgage market that began in the United States in fall 2006 and became to a global financial crisis in July 2007.

In the case of Germany, the collapse of Lehman Brothers on September 15, 2008, marked the climax of the financial crisis and the beginning of the economic recession, as it was announced on November 13, 2008, by Germany's Federal Statistical Office. The name of the crisis stems from a special segment of the U.S. mortgage market, the market for subprime loans. Basically, mortgage companies and banks were lending too much money to people who could not afford the houses they were purchasing. Germany's banks, already heavily exposed to subprime securities, have been particularly affected by the acute, ongoing tension in the money markets and the financial markets, which have been devastated by a massive flight to quality. Germany's highly export dependent economy is losing steam following a dramatic slowdown in its overseas markets.

In the case of Germany, the stock market declines are largely attributable to steep losses in the financial industry in the year 2008, as for example, the share price of Commerzbank with losses of 52,9 percent, with Deutsche Bank declining by 26,65 percent, Deutsche Postbank falling by 45,08 percent, Infineon by 25,87 percent and Siemens by 18,17 percent. But more than anything, the skittishness on the stock market in Germany is being caused by the Lehman bankruptcy.

The first important tremors in the German stock market during the subprime crisis in 2008 were on January 21, the Global stock markets, including London's FTSE 100 index, suffered their highest declines since 11 September 2001.

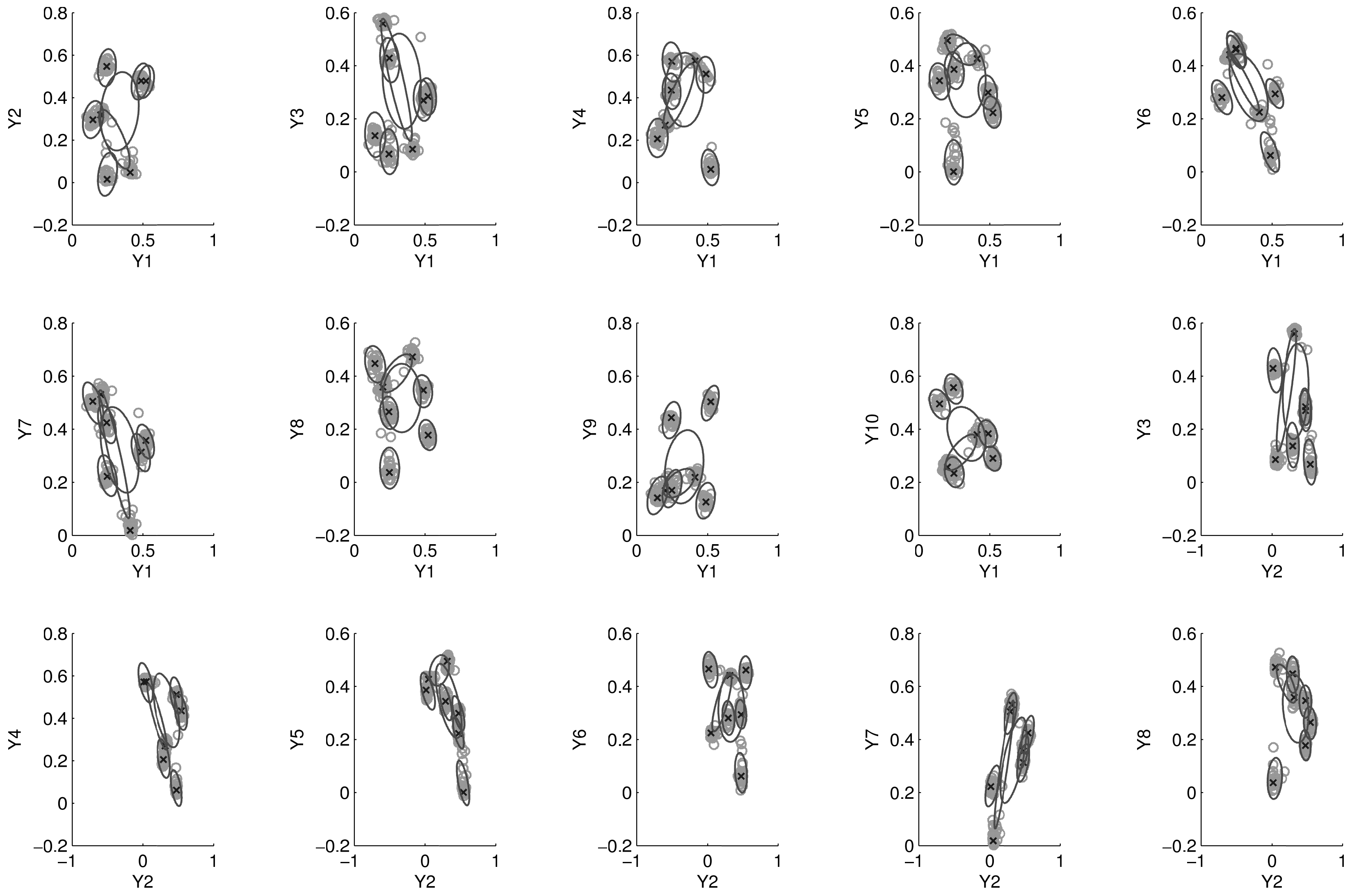

FIGURE 5.6.2. Bivariate projections of a DPM. The $\times$ in black colours are the true signature patterns $Y_p = \{y_{1p}, \dots y_{l'p}\}$, while the points in grey colours are the empirical estimations of them. The ellipsoids displayed in the figures are given by the eigenvectors of the estimated covariance matrix Ω.

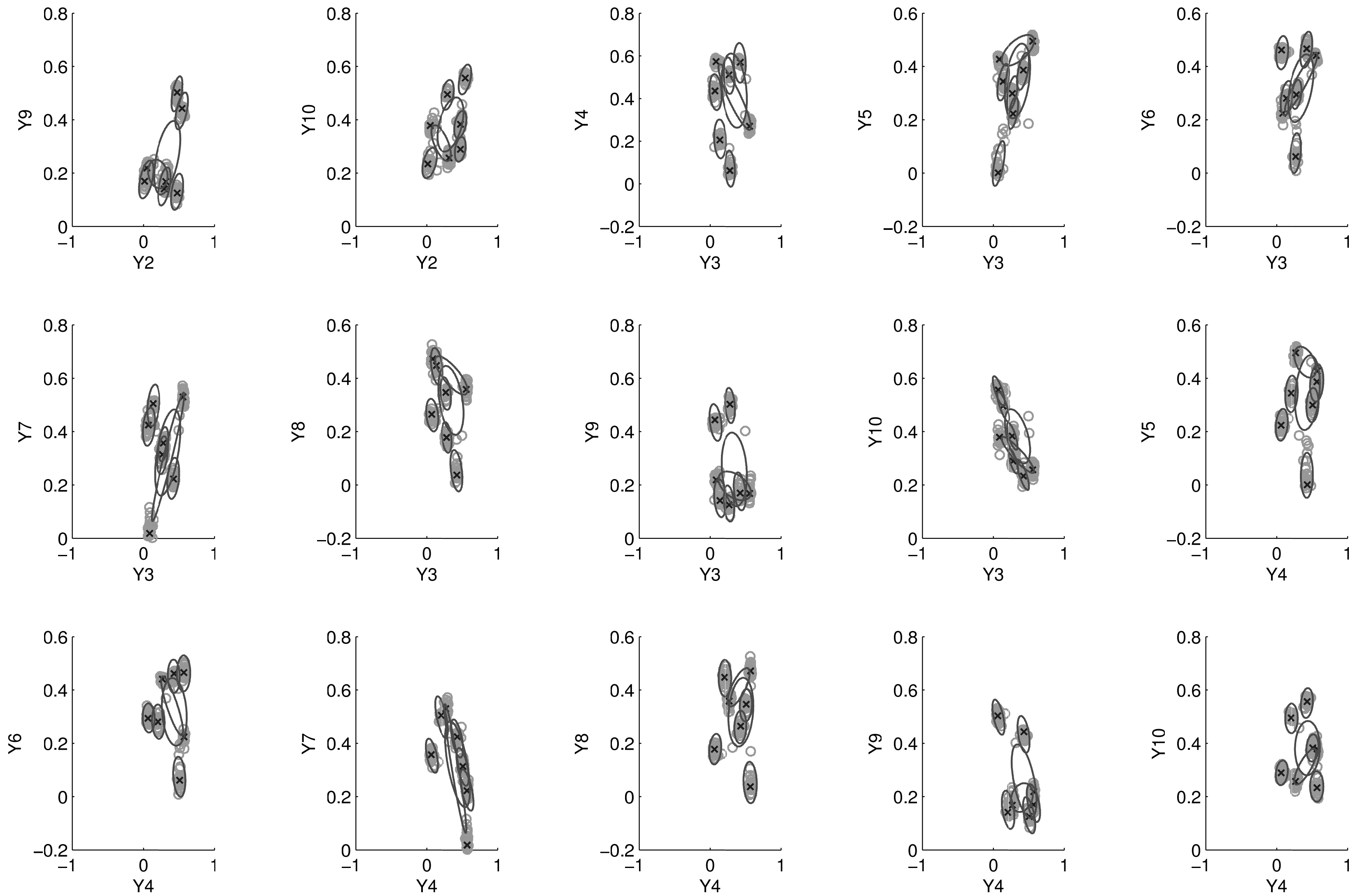

FIGURE 5.6.3. (Cont.) Bivariate projections of a DPM. The $\times$ in black colours are the true signature patterns $Y_p = \{y_{1p}, \dots y_{l'p}\}$, while the points in grey colours are the empirical estimations of them. The ellipsoids displayed in the figures are given by the eigenvectors of the estimated covariance matrix Ω.

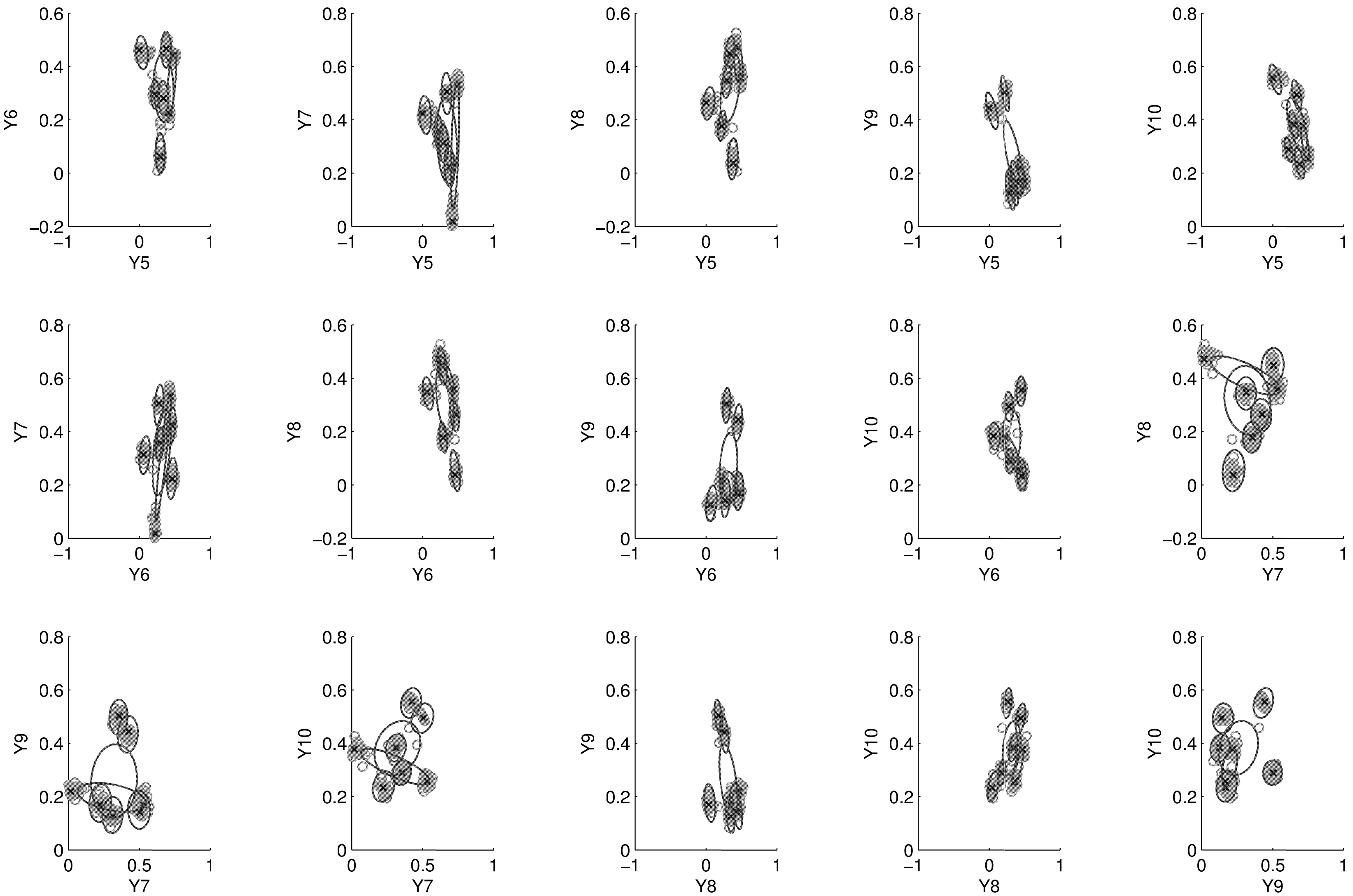

FIGURE 5.6.4. (Cont.) Bivariate projections of a DPM. The $\times$ in black colours are the true signature patterns $Y_p = \{y_{1p}, \ldots y_{l'p}\}$, while the points in grey colours are the empirical estimations of them. The ellipsoids displayed in the figures are given by the eigenvectors of the estimated covariance matrix Ω.

Company	Industry Group	Symbol	Index weighting (%)
E.ON	multi-utilities	EOA	10.00
Siemens	diversified industrials	SIE	9.83
Allianz	insurance	ALV	7.34
Volkswagen Group	automobile manufacturers	VOW	7.28
Bayer	speciality chemicals	BAY	7.06
RWE	multi-utilities	RWE	5.73
BASF	speciality chemicals	BAS	5.62

TABLE 5.7.1. Market participation of some of the major current companies present in the DAX index.

The DAX Index has tumbled 45 percent since the beginning of last year as credit losses and writedowns topped \$1 trillion in the worst financial crisis since the Great Depression and together with countries like the U.S., Japan and Europe fell into simultaneous recessions.

Every market crisis brings significant losses to the overall economy and its impact should be minimized as much as possible. This section is motivated by the concrete question: can we give, in an objective and non arbitrary manner, good approaches to extreme risk measures in a multivariate frame work?

We provide an answer to this question through the application of M_4 processes on a sector of the DAX index companies consisting of the 7 of the major German companies trading on the Frankfurt Stock Exchange in terms of order book volume and market capitalization. These are listed in Table 5.7.1.

The data set consists of daily returns defined by $r_t = -100 \ln(p_t/p_{t-1})$, where p_t denotes the value of the index at day t, from the stock market indices over a sample period from 1 January 1973 to 19 January 2008, one day before on January 20 the Global stock markets suffered their biggest falls since 11 September 2001.

A second sample is used for backtesting the estimation of the different risk measures in the DAX index from January 20, 2008 to February 26, 2009. In order to better understand the empirical exercise, it is worth looking briefly at the basic characteristics of the analyzed financial series. in Table 6.B.3 in the Appendix, we find some descriptive statistics of the daily returns on the above series. The mean return is close to zero for all of the seven series. However, it differs considerably in terms of standard deviation, skewness and kurtosis of a normally distributed random variable.

The assumption of normally distributed returns is strongly rejected by our series through the Jarque-Bera test. The high value for excess kurtosis indicates that the distributions are characterized by leptokurtosis. Moreover, we compute the augmented Dickey-Fuller test for the null hypothesis that the returns have a unit root. This hypothesis is also rejected for the whole series.

5.7.1. Approximation by M_4 processes. In general, M4 process models cannot be directly fitted to observed data since the data is not in unit Frèchet scales. Certain scale transformation is needed. Our strategy to analyze returns is a three-stage approach.

First fit a volatility model to remove the mean and volatility effects. Second, fit an extreme value model to the standardized residuals from the first stage and transform to unit

Frèchet scale. Third, fit an M_4 process model to the variables in the unit Frèchet scale. Our model takes into account several empirical facts of financial time series, such as time-varying volatility, heavy tails, and extreme dependence across assets.

First stage: filtering volatility. In the first stage of modeling we fit GARCH(1,1) models to each of the individual returns in order to remove trend and volatility effects. The GARCH estimations are depicted in Table 6.B.4. Then we calculate standardized residuals. The negative log returns, the estimated standard deviations and the standardized residuals are drawn in Figure 5.B.1. From the plots, clearly we see that there are extreme observations, jumps in returns, jumps in volatilities in each sequence.

One can observe that the devolatilized time series now looks stationary. However, jumps in returns are still persistent.

Second stage: transformation to Frèchet scale. The next step is to fit a generalized Pareto distribution (GPD) to exceedances over a specified threshold value, in our case the 0.90 th-quantile for all returns. Let us focus on a standardized residual Z. The distribution function for the exceedances of a threshold u by the variable Z, conditional in $Z > u$ for large enough u, is given by

$$G(z) = 1 - \lambda \left(1 + \xi \frac{(z-u)}{\sigma}\right)_+^{-1/\xi} \tag{5.7.1}$$

where (λ, σ, ξ) are the proportion of observation over the threshold u ($\lambda = \mathbb{P}(Z > u)$), the scale and shape parameters respectively and $h_+ = \max(h, 0)$.

We report the estimations in Table 5.B.3. Figure 5.B.2 gives detailed information of the fit and model diagnostic tools to the real data. The plots compare the parametric distribution, densities, and quantiles to their empirical counterparts (see Coles (2001) for a detailled description). The GPD fitting is applied to standardized positive returns (in our case the negative returns). These plots indicate good fits for all seven variables.

Using the fitted GPD distributions, the residuals data is transformed into unit Frèchet scale as follows.

$$X = -\left(\log\left(G(z)\right)\right)^{-1}, \tag{5.7.2}$$

for observations over the threshold u, while we use a simple ranking transformation (see Heffernan and Resnick (2005)) for the observations under this threshold. We apply analogously this transformation to each marginal. These final transformed data is the base of M_4 process modelling.

The purpose here is to model those values above the threshold value by M_4 models, therefore, a logic way is to chose the $d-$dimensional returns where at least a extreme event has been observed, i.e., $\boldsymbol{X}_i \nleq \boldsymbol{u}$ or equivalently $X_{i,j} > u_j$ for some $j = 1, \ldots, d$.

Thrid stage: The M_4 process estimation. For M_4 processes, we will analyse the temporal dependence between exceedances (the cluster size coefficient $\mathcal{K}$) over threshold sequences by means of the extremal index and the cluster size probabilities as was introduved in Ferro and Segers (2003) and chapter 3.

We obtain estimations of the extremal index and cluster size probabilities by these methods. Table 5.B.4 shows that these estimations fluctuate around 0.52 and 0.63, so the mean cluster size of exceedances is in the interval (1.59, 1.88), which indicates that there is clustering of extreme events that is not captured by the volatility model. The cluster probability indicates that most of the extreme events happen in cluster size of only 1 or 2 components. Based on this analysis, we propose to work with a cluster size $\mathcal{K} = 2$.

We specify an M_4 process model in seven dimensions and $\mathcal{K} = 2$, this is $k = 0, 1$ and we fit this model using the Dirichlet process mixture described in Algorithm 5.5.1. Table 5.B.5 shows the parameter estimation of the M_4 process model fitted to the losses of the German stock markets returns in the Frèchet scale. Remark that actually the estimations are the mean of our infinite Gaussian mixture together with the standard deviation obtained from the covariance matrix Ω. We report only the standard deviation by the fact that the related covariance are six matrices of 14×14. The concentration parameter $\alpha_0 = 0.9746$. The standardized results for the coefficients a_{jil} are presented in Table 5.B.6. These were calculated as was described in subection 5.2.1. The cluster size n_l for each signature pattern were 69, 36, 52, 63, 56, and 57 respectively.

We simulate a sample of size 20.000 to test if the fitted M4 process is a realistic representation of the Frèchet-transformed time series and if it can be utilized in common measures in risk management. In a first look at the quality of the simulated data, we displayed a scatterplot among the sample paths simulated from the fitted process and observed if this sample looks similar to those from the original series. Figure 5.7.1 in background and with grey colours is the simulated sample, while the original sample is in black colour. One point to note here is that the data was generated so that the marginal distributions were exactly unit Frèchet. In order to provide a fair comparison with the original estimation procedure, the axes were rescaled so that a considerable number of points were present. The results are optimistic, due to the fact that the cluster and direction of the extremes seems to give a good approach to the real data.

The simulated M_4 process is used to obtain estimations of the extremal indices of the individual series. Notice that with equation (5.2.5) the multivariate extremal index could also be calculated but it depends on the threshold chosen.

The estimations are depicted in Table 5.B.7. The results indicate that the simple approach mimics the cluster behaviour at the extremes of the original serie. However, in a most deeper analysis we observed that the cluster size probabilities are different. In fact the estimated process tends to favour the cluster of size two. This is not a surprise by the simple representation in the size cluster coefficient k that we have chosen. In the next section we give an application based on the model estimated.

5.7.2. Market risk and stress testing with M_4 processes. The Basel I and II Accord requires regulatory capital to cover market risk and stress tests in a coherent and objective framework. In this section we focus on particular on the process of stress testing from identifying vulnerabilities and predictability for the German stock market, to constructing scenarios and backtesting the results.

We propose the M_4 process in this context which can incorporate both clustering and heavy tails. In a first approach we illustrate an application to portfolio optimization under VaR constraints. In a second application we establish different scenarios to stress test different

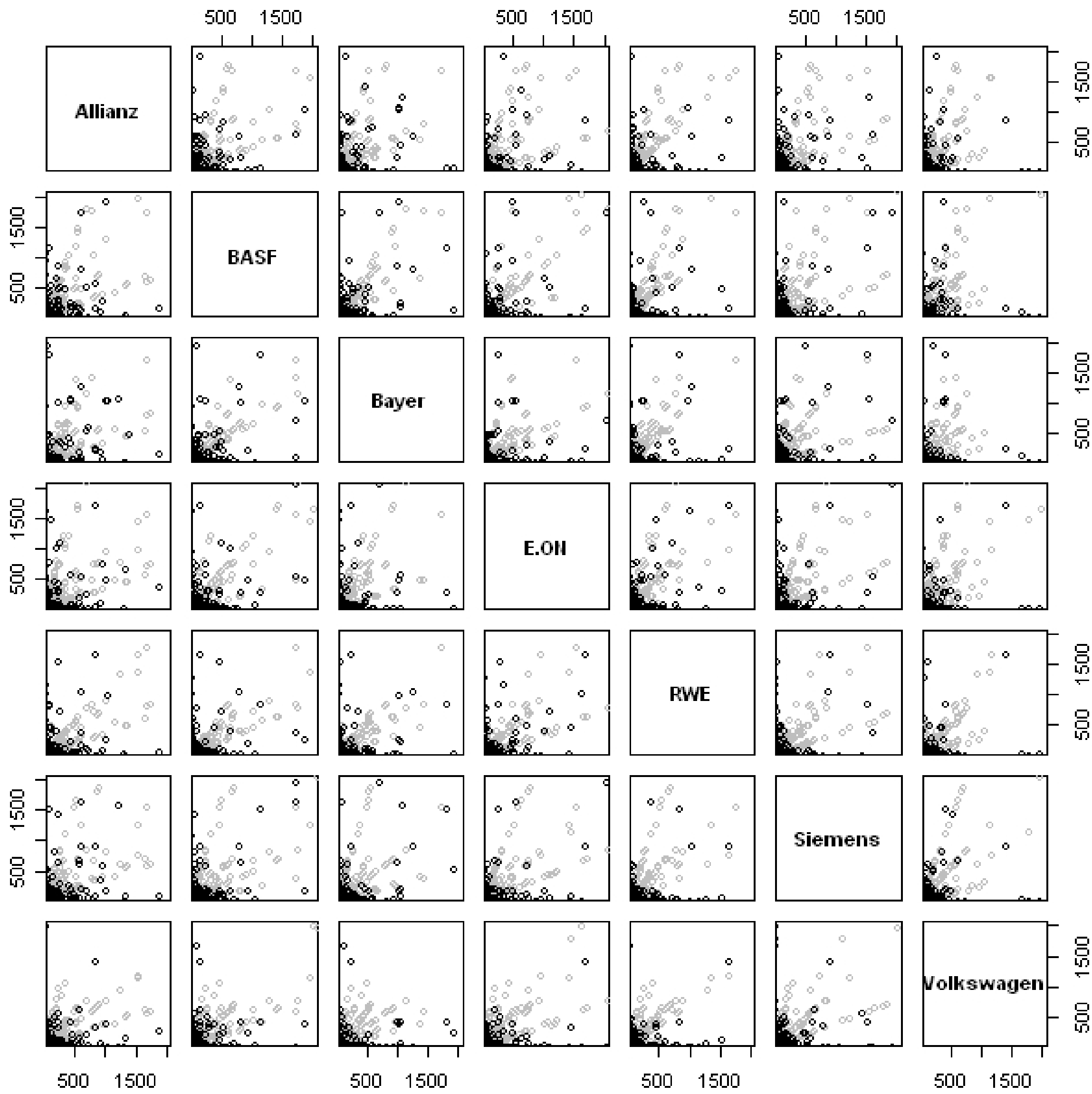

FIGURE 5.7.1. Scatterplot of standardized DAX returns (in black colour) and the estimated M_4 process (in grey colour) in Frèchet scale. The size sample simulation is 20.000.

portfolios obtained in the first application. Traditional stress testing are based on historical data or they are hypothetical and can involve large movements, hence the probability of an extreme outcome is unknown and many extreme yet plausible possibilities are ignored. Many stress tests also fail to incorporate the characteristics that markets are known to exhibit in crisis periods, namely, increased probability of further large movements, increased comovement between markets, greater implied volatility and reduced liquidity. We make use of the second part of the sample from January 20, 2008 to February 26, 2009 to backtest the results.

The first problem presents in the estimation of extreme scenarios is that the VaR for the worst case, in multidimensional case, has been recent studied in the literature and only

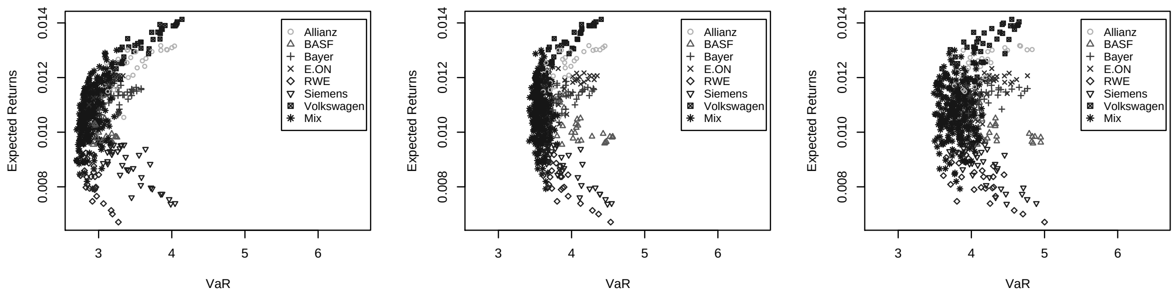

FIGURE 5.7.2. VaRs for portfolios of different combinations of the seven major German stock market returns. All VaRs are calculated based on the standardized data. The three approaches are (left) multivariate normal, (middle) historical simulation and (rigth) M_4 processes.

for the 2- dimensional case some bounds are known (see Embrechts et al. (2005)). For this reason, we adopted the following framework proposed by Zhang (2002). Suppose $w_1, \ldots, w_7$ are proportions of the German stock returns in a portfolio, VaR^d is the VaR of the portfolio return for a given level α as for example $\mathbb{P}\left(\sum_{j=1}^{d} w_j X_j > VaR^d\right) < \alpha$.

The worst case scenario for the VaR^d for a level α is determined in the case of M_4 processes simultaneously by the next optimization problem

$$\begin{aligned} \arg\max \qquad & \mathbb{P}\left(\bigcap_{j=1}^{d} w_j X_j > VaR_j\right) \\ \text{subject to} \quad & \mathbb{P}\left(\sum_{j=1}^{d} w_j X_j > VaR^d\right) < \alpha \\ & \sum_{j=1}^{d} VaR_j = VaR^d \end{aligned}$$

where VaR_j is the individual value at risk of the factors. The objective is to have the highest probability for all individual risk factors beyond certain values when the portfolio is at the worst case of VaR^d. The factors VaR_j are determined inversely by equations (5.7.1) and (5.7.2).

In the literature there are two typical methods to calculate VaR, variance-covariance approach and historical simulation approach. The variance-covariance approach assumes the return has a normal distribution for a single asset or the returns have jointly multivariate normal distribution for a portfolio with multiple assets, while the historical simulation is based on extreme events of the past (see for example Markowitz (1991) and Jorion (2003) for more references).

We simulated 500 strategies for the German stock markets and calculate the mean-VaR portfolio at the $1 - \alpha = 0.99$ level of confidence for each strategy of the standardized data. The market is assumed to be arbitrage free and without friction. The weights correspond to w_j for each stock market and the sum of them is equal to one. Figure 5.7.2 displayed the results for the three simulations. Different symbols and colours are utilized in connection with the stock markets to denote that more than half of the weights are designed to a specific class of stock market. The *Mix* strategy refers to a portfolio composition, which contains less that 50% of the total shares of some class of stock markets in a portfolio.

In Figure 5.7.2 each of the three methods suggests that the risk can be reduced if a diversification of investment is applied, i.e, the *Mix* strategy. Portfolios where the most influential stock markets are Allianz and Volkswagen give the highest VaR together with highest expected return. In the variance-covariance approach stock Siemens has the lowest expected return together with the highest VaR, while in the historical simulation and in the M_4 approach case Siemens gives the highest VaR. Comparing the three methods the variance-covariance approach always gives lower estimated VaR. The VaRs computed from the historical simulation and the M_4 approach are similar but with some higher VaR estimations for strategies where a component contains more than 50% of the total shares of some class of stock markets in the portfolio.

Stress testing and back testing. The next step is to considerer different scenarios based on the three approaches in portfolio optimization to backtest one year of the sample, from January 20 to February 26, 2009. Stress testing techniques fall into two general categories: sensitivity tests and scenario tests. Sensitivity tests assess the impact of large movements in financial variables on portfolio values without specifying the reasons for such movements. A typical example might be a 10% decline in some stock market index as the Volkswagen. These tests can be run relatively quickly and are commonly used as a first approximation of the portfolio impact of a financial market move. However, the infinite number of scenarios, the analysis lacks historical and economic content can limit its usefulness for longer term risk-management decisions.

On the other hand, the introduction of subjective scenarios with assigned probabilities should allow us to create a more comprehensive picture of risk including all available information. These stress tests are typically applied at a point in time or in conjunction with a forecast over a specific horizon. For example, assuming only a limited behavioural response in a large portfolio over a one to three month horizon, because it is often difficult to restructure a portfolio in less time without incurring losses from "fire-sale" prices. But once the time horizon of a scenario extends beyond a year or more, the assumption of no feedback effects implicit in many stress tests may be an oversimplification.

In the present study the "perfect storm" happened. For 6 of the 7 stock markets considered the losses were in the last year in the order of -38% Allianz, -29% BASF, -19.5% Bayer, -33.5% E.ON, -24.5% RWE and -31.6% Siemens[1]. The only stock without losses was Volkswagen with a 4.4% gain.

How to obtain good results in these extreme conditions? Regardless of the oversimplification, we chose 20 of the 500 portfolios combinations for backtesting one year of the sample for each method of portfolio optimization approach. We take the 20 more rentable portfolios in terms of the expected return, whose VaR at the 0.99-th quantile is under the 25% of the riskier portfolios estimated through the Montecarlo simulation realized. The results of this backtest for a long position are depicted in Figure 5.7.3. On the left of the figure we have

[1]These losses are calculated from the standarized data, i.e., after the GARCH filter.

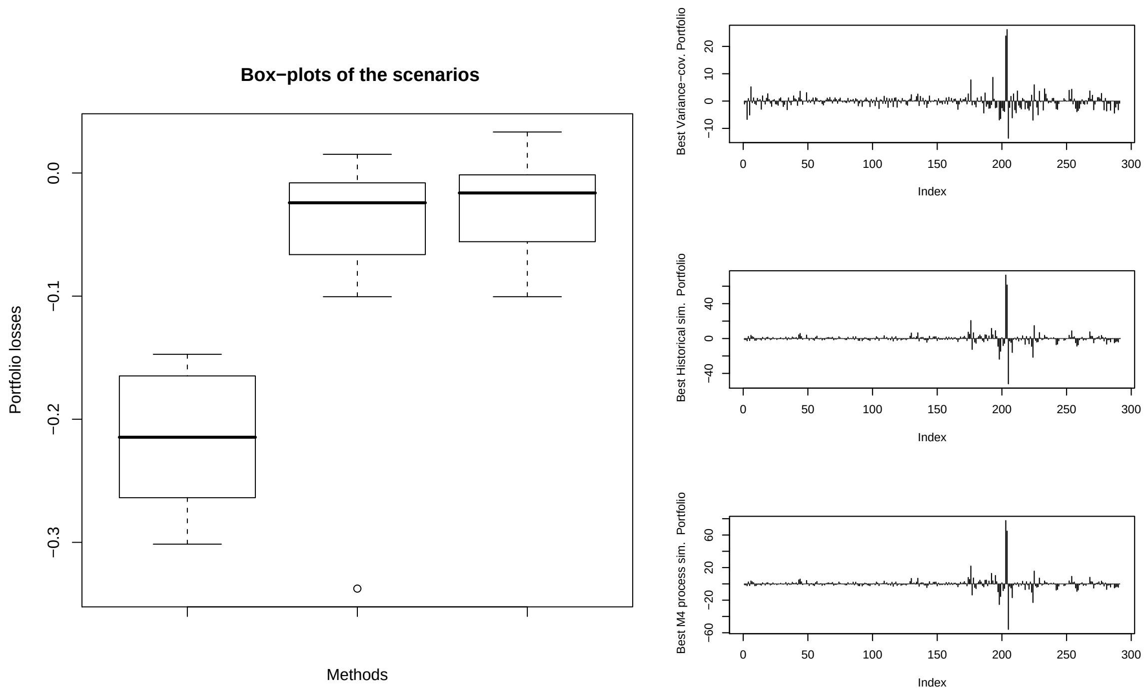

FIGURE 5.7.3. (Left panel) Box-plots of the performance of the 20 scenarios choice for each portfolio given by (right) variance-covariance method, (middle) historical simulation and (right) M_4 processes. (Right panel), Evolution of the returns for the best porfolio perfomance for each approach, from top to bottom; variance-covariance method, historical simulation and M_4 processes.

a boxplot (also known as a box-and-whisker diagram or plot) of the performance of the 20 scenarios choice for each portfolio optimization method. It is a convenient way of graphically depicting of numerical performance of the portfolios. The spacings between the different parts of the box help indicate the degree of dispersion and skewness in the data, and identify outliers. The left boxplot corresponds to the variance-covariance method, in the middle the historical simulation and to the right the M_4 approach. We observe that the best performance is given by the M_4 approach followed by the historical simulation. The variance-covariance approach underestimates the losses completely. The most of the portfolios get losses under the assumption of no feedback to re-estimate the models. Nevertheless, what the M_4 approach is doing, is making the best use of whatever data you have about extreme events and limiting their influence in the performance of a portfolio. In the right side of the Figure 5.7.3 we illustrate the evolution of the returns for the best porfolio perfomance for each approach. The M_4 approach war the only one without losses during the backtesting.

In resume, the results of the approaches provide a useful check of the validity of M_4 processes as framework. The variance-covariance approach fails basically by the multivariate Gaussian assumption for the distribution (light tailed distribution) of the stock market returns and the implicit use of the covariance matrix as measure of dependence among other things. Historical scenarios can be more intuitive because they were actually observed, but the scenarios given by the M_4 approach are more realistic in case of the recent crisis, especially if the financial structure has changed significantly.

5.8. Conclusions

The rescent international financial turmoil has prompted the development of new frameworks, risk tools, and techniques to assess the stability of financial systems. But what happens if an actual shock hits the system - how will the system perform? Multivariate extreme value theory can be used as an additional instrument in the financial sector to help answer such questions.

In this chapter we concentrate on the estimation of a class of max-stable processes, M_4 processes. The main advantage of this class of processes is that its representation allows us to approximate multivariate time series with cluster at the extremes and heavy tail behaviour. We proposed a nonparametric Bayesian approximation to the estimation of M_4 processes. The idea was to estimate the singularities in the multivariate density, characteristics of this process, through a infinite mixture of Dirichlet processes of Gaussian distributions. These singularities have a cluster behaviour, which repeat infinitely, hence it can be observed. We solve the problem of estimated the number and behaviour of these singularities assuming that these behave as an infinite multivariate Gaussian mixture, where a finite number is inferred from the data observed available.

From the results in the experimets we can conclude that DP mixtures can give an accuracy estimation of the signature patterns in a M_4 process. We observe further, that the estimation of the exact number of patterns $\mathcal{L}$ in a M_4 process depend more in the number of exceedances over a threshold $\boldsymbol{u}$, while the number of dimensions play a second role.

We propose the M_4 process in this context of risk management incorporating both clustering and heavy tails to the analysis of some selected German stock markets. In a first approach we illustrate an application to portfolio optimization under VaR constraints. We compare M_4 process with other two approach variance-covariance and historical simulations. The VaRs computed from the M_4 approach are higher in general. Moreover, the results indicate that the M_4 approach mimics the cluster behaviour at the extremes of the original serie.

In a second application, we establish different scenarios to stress test different portfolios obtained in the first application. The scenarios given by the M_4 approach are more realistic in case as the recent crisis. The best hedging is also given by this approach.

In relation to the implementation, Monte carlo markov chains methods, especially Gibbs sampling is straightforward and accurate. However, they can be prohibitively slow, especially in the context multidimensional as our case. The most large M_4 model in this chapter was of 15 dimensions with cluster size $\mathcal{K} = 2$, which means in our framework 30 dimensions. This take about 20 minute to be solved in a ThinkPad 60 Core2Duo and 2G RAM. One would like to be able to consider alternative inference algorithms for the DP. In this sense in future works we will concentrate on variational methods (Blei and Jordan (2006); Kurihara et al. (2007)), which have provided fast deterministic alternatives to MCMC for approximating otherwise intractable posteriors in simpler settings.

Future works might concentrate on adapting the framework to variational methods. Other idea could be explote a hierarchical DP structure (Teh et al. (2006)) in relation to the cluster size $\mathcal{K}$ to reduce the number of dimensions.

5.A. Demonstrations

PROOF. Lemma 5.2.5: For $l \in \mathbb{N}$ and $r \in \mathbb{Z}$, let $\breve{a}_{l,r} = \max_{j=1,\dots d} a_{lrj}/u_j$. Then

$$\begin{aligned}
W_4(\boldsymbol{u}, n) - W_4(\boldsymbol{u}, n-1) &= \sum_{l \in \mathbb{N}} \sum_{k \in \mathbb{Z}} \left(\max_{i=1,\dots,n} \breve{a}_{l,i-k} - \max_{i=1,\dots,n-1} \breve{a}_{l,i-k} \right) \\
&= \sum_{l \in \mathbb{N}} \sum_{k \in \mathbb{Z}} \left(\breve{a}_{l,n-k} - \max_{i=1,\dots,n-1} \breve{a}_{l,i-k} \right)_+ \\
&= \sum_{l \in \mathbb{N}} \sum_{r \in \mathbb{Z}} \left(\breve{a}_{l,r} - \max_{i=1,\dots,n-1} \breve{a}_{l,i+r-n} \right)_+ \\
&= \sum_{l \in \mathbb{N}} \sum_{r \in \mathbb{Z}} \left(\breve{a}_{l,r} - \max_{i=1,\dots,n-1} \breve{a}_{l,r-i} \right)_+
\end{aligned}$$

where $\alpha_+ = \max(\alpha, 0)$ for $\alpha \in \mathbb{R}$. By the dominated convergence theorem,

$$\begin{aligned}
\lim_{n \to \infty} (W_4(\boldsymbol{u}, n) - W_4(\boldsymbol{u}, n-1)) &= \sum_{l \in \mathbb{N}} \sum_{r \in \mathbb{Z}} \left(\breve{a}_{l,r} - \max_{s<r} \breve{a}_{l,s} \right)_+ \\
&= \sum_{l \in \mathbb{N}} \max_{r \in \mathbb{Z}} \breve{a}_{l,r} \\
&= W_4'(\boldsymbol{u})
\end{aligned}$$

Moreover, since the Cesàro transformation of a converging sequences converges to the same limit as the original sequence, also

$$\begin{aligned}
\lim_{n \to \infty} W_4(\boldsymbol{u}n, n) &= \lim_{n \to \infty} \frac{1}{n} W_4(\boldsymbol{u}n, n) \\
&= \lim_{n \to \infty} \frac{1}{n} \sum_{r=1}^{n} (W_4(\boldsymbol{u}, k) - W_4(\boldsymbol{u}, k-1)) \\
&= W_4'(\boldsymbol{u})
\end{aligned}$$

This conludes the proof. □

PROOF. Proposition 5.2.6: Let $\mathbf{X}$ be the bivariate process, then

$$\begin{aligned}
\lambda &= \lim_{u_1, u_2 \to \infty} \mathbb{P}\left(X_1 > u_1,\ X_2 > u_1\right) / \mathbb{P}\left(X_2 > u_1\right) \\
&= \lim_{u_1, u_2 \to \infty} \frac{1 - M_4\left(u_1\right) - M_4\left(u_2\right) + M_4\left(u_1, u_2\right)}{1 - M_4\left(u_2\right)} \\
&= \lim_{u_1, u_2 \to \infty} \frac{1 - \exp\left(-1/u_1\right) - \exp\left(-1/u_2\right) + \exp\left(- \sum_{l \geq 1} \sum_{k \in \mathbb{Z}} \max_{1 \leq j \leq 2} \left\{a_{lkj}/u_j\right\} \right)}{1 - \exp\left(-1/u_2\right)}.
\end{aligned}$$

Using the asymptotic identity $\lim_{u \to \infty} \exp\left(-1/u\right) \sim \lim_{x \uparrow 1} 1 - \log\left(x\right)$, we get

$$\begin{aligned}
\lambda &= \lim_{x \uparrow 1} \frac{1 - (1 - \log x) - (1 - \log x) + \left(1 - \log x \sum_{l \geq 1} \sum_{k \in \mathbb{Z}} \max_{1 \leq j \leq 2} \left\{a_{lkj}\right\} \right)}{\log x} \\
&= 2 - \sum_{l \in \mathbb{N}} \sum_{k \in \mathbb{Z}} \max_{1 \leq j \leq 2} \left\{a_{lkj}\right\}
\end{aligned}$$

□

The same result can be obtained by using the L'Hopital rule.

Proof. Proposition 5.2.10: Observe that for example for the marginal 1

$$a^{\Sigma}_{l*}\frac{X_{i*+k,l}}{\sum_{\mathcal{K}} X_{i*+k}} = \frac{a_{l^*k}Z_{l^*,i*} + N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*} + N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}$$

which is approximately equal to

$$\begin{aligned}
a^{\Sigma}_{l*}\frac{X_{i*+k,1}}{\sum_{\mathcal{K}} X_{i*+k}} &= \frac{\frac{a_{l^*k}}{\sum_{\mathcal{K}} a_{l^*k}} + \frac{N_{1(i*+k)}\sqrt{1-\rho^2}+\rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}}}{1 + \frac{N_{1(i*+p)}\sqrt{1-\rho^2}+\rho N_{2(i*+p)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}}} \\
&\sim \left(\frac{a_{l^*k}}{\sum_{\mathcal{K}} a_{l^*k}} + \frac{N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}}\right) \times \\
&\quad \left(1 - 1 + \frac{N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}}\right) \\
&= \frac{a_{l^*k}}{\sum_{\mathcal{K}} a_{l^*k}} + \frac{N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}} \\
&\quad - \frac{a_{l^*k}}{\sum_{\mathcal{K}} a_{l^*k}} \times \frac{N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}} \\
&\quad - \left(\frac{N_{1(i*+k)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}}{\sum_{\mathcal{K}} a_{l^*k}Z_{l^*,i*}}\right)^2
\end{aligned}$$

Since the tails of $Z_{l^*,i*}$ are larger in comparison to $N_{1(i*+pk)}\sqrt{1-\rho^2} + \rho N_{2(i*+k)}$ we get

$$a^{\Sigma}_{l*}E\left(\frac{X_{i*+k,1}}{\sum_{\mathcal{K}} X_{i*+k}, 1}\right) \sim \frac{a_{l^*k}}{\sum_{\mathcal{K}} a_{l^*k}}.$$

□

5.B. Tables and Figures

	Allianz	BASF	Bayer	E.ON	RWE	Siemens	Volkswagen
N° obs.	9433	9433	9433	9433	9433	9433	9433
Mean	0.022	0.019	0.020	0.024	0.023	0.019	0.036
Standard deviation	1.864	1.492	1.604	1.531	1.476	1.671	2.293
Minimum	-15.678	-12.924	-18.432	-13.977	-15.822	-16.364	-58.044
Maximum	19.273	12.691	32.310	15.886	14.255	16.602	80.528
Skewness	0.274	-0.228	0.486	-0.114	0.137	-0.170	5.145
Kurtosis	11.761	6.911	24.062	8.207	8.387	9.828	287.055
Ljung-Box test	40.867*	16.265*	11.442*	67.043*	27.974*	44.889*	314.551*
Jarque-Bera test	54517*	18865*	228035*	26508*	27692*	38028*	32442379*
Augmented Dickey-Fuller test	-19.905	-21.517	-20.609	-21.057	-22.162	-20.288	-19.475

TABLE 5.B.1. Summary statistics for the stock market returns. Asymptotic p-value are shown in the brackets. *,**,*** denote statistical significance at the 1, 5 and 10 % level respectively. The Ljung-Box test statistic for serial correlation up to the 5-th order.

	Allianz		BASF		Bayer		E.ON		RWE		Siemens		Volkswagen	
	coeff.	s.e	coeff.	s.e	coeff.	s.e	coeff.	s.e	coeff.	s.e	coeff.	s.e	coeff.	s.e
μ	-0.046	0.013	-0.037	0.012	-0.037	0.013	-0.046	0.013	-0.037	0.011	-0.031	0.011	-0.048	0.016
ω	0.042	0.005	0.090	0.010	0.042	0.006	0.060	0.008	0.020	0.004	0.009	0.002	0.078	0.009
α	0.112	0.008	0.113	0.010	0.079	0.007	0.081	0.008	0.059	0.006	0.050	0.005	0.095	0.006
β	0.883	0.008	0.847	0.012	0.906	0.008	0.893	0.010	0.933	0.007	0.948	0.005	0.888	0.007
Log Lik.	17164.60		16030.15		16433.84		16218.33		15652.76		15971.99		18436.39	

TABLE 5.B.2. GARCH(1,1) estimates with Normal distributed noise. All the coefficients are significative at the 0.01%.

	scale σ		shape ξ	
	Estimate	s.e	Estimate	s.e
Allianz	1.2119	0.06258	0.1998	0.04061
BASF	0.9651	0.04923	0.1760	0.03969
Bayer	1.0623	0.05279	0.1537	0.03788
E.ON	1.0693	0.05176	0.1355	0.03610
RWE	1.00805	0.04806	0.09949	0.03498
Siemens	1.1711	0.05872	0.1585	0.03848
Volkswagen	1.0491	0.05251	0.2797	0.03911

TABLE 5.B.3. Parameter Estimates of GPD Models with a 0.9-th quantile threshold for all resturns.

Stock markets	extremal index	$\pi(1)$	$\pi(2)$	$\pi(3)$	$\pi(\leq 4)$
Allianz	0.531	0.615	0.253	0.080	0.052
BASF	0.596	0.643	0.242	0.073	0.043
Bayer	0.574	0.640	0.264	0.057	0.039
E.ON	0.626	0.668	0.234	0.076	0.022
RWE	0.578	0.608	0.269	0.088	0.035
Siemens	0.542	0.643	0.246	0.066	0.045
Volkswagen	0.530	0.604	0.272	0.088	0.036

TABLE 5.B.4. Extremal indices and cluster size probabilities $\pi(k)$ of size k.

	Allianz		BASF		Bayer		E.ON		RWE		Siemens		Volkswagen	
l	$a_{l,0,1}$	$a_{l,1,1}$	$a_{l,0,2}$	$a_{l,1,2}$	$a_{l,0,3}$	$a_{l,1,3}$	$a_{l,0,4}$	$a_{l,1,4}$	$a_{l,0,4}$	$a_{l,1,4}$	$a_{l,0,4}$	$a_{l,1,4}$	$a_{l,0,4}$	$a_{l,1,4}$
1	0.074	0.179	0.191	0.073	0.123	0.059	0.225	0.061	0.084	0.065	0.091	0.075	0.083	0.067
2	0.032	0.033	0.027	0.027	0.032	0.032	0.029	0.022	0.142	0.138	0.039	0.033	0.034	0.027
3	0.069	0.042	0.068	0.045	0.071	0.058	0.084	0.160	0.071	0.047	0.219	0.062	0.084	0.050
4	0.204	0.104	0.072	0.083	0.096	0.065	0.086	0.091	0.097	0.082	0.070	0.217	0.068	0.082
5	0.059	0.089	0.074	0.197	0.157	0.161	0.054	0.055	0.051	0.095	0.047	0.043	0.043	0.067
6	0.046	0.069	0.047	0.094	0.047	0.100	0.043	0.089	0.048	0.080	0.042	0.062	0.191	0.206

TABLE 5.B.6. Coefficients a_{lkd} of the fitted M_4 process to the German stock markets. These were calculated by the procedure proposed in section 5.2.

Stock markets	extremal index	$\pi(1)$	$\pi(2)$	$\pi(3)$	$\pi(\leq 4)$
Allianz	0.503	0.357	0.485	0.103	0.055
BASF	0.521	0.421	0.434	0.090	0.055
Bayer	0.468	0.276	0.572	0.076	0.077
E.ON	0.506	0.396	0.459	0.083	0.061
RWE	0.459	0.215	0.625	0.081	0.080
Siemens	0.529	0.435	0.421	0.090	0.055
Volkswagen	0.457	0.186	0.659	0.075	0.080

TABLE 5.B.7. Extremal indices and cluster size probabilities $\pi(\cdot)$.

$\boldsymbol{\mu}_{lp}$	μ_{l1}	μ_{l2}	μ_{l3}	μ_{l4}	μ_{l5}	μ_{l6}	μ_{l7}	μ_{l8}	μ_{l9}	μ_{l10}	μ_{l11}	μ_{l12}	μ_{l3}	μ_{l14}
μ_{1p}	0.120	0.289	0.308	0.118	0.189	0.090	0.355	0.096	0.135	0.105	0.150	0.122	0.137	0.111
	(0.143)	(0.252)	(0.256)	(0.125)	(0.200)	(0.112)	(0.299)	(0.111)	(0.143)	(0.121)	(0.183)	(0.146)	(0.179)	(0.148)
μ_{2p}	0.098	0.101	0.083	0.084	0.094	0.095	0.089	0.067	0.438	0.427	0.124	0.102	0.107	0.085
	(0.137)	(0.142)	(0.118)	(0.122)	(0.129)	(0.145)	(0.140)	(0.120)	(0.289)	(0.314)	(0.177)	(0.153)	(0.129)	(0.144)
μ_{3p}	0.148	0.089	0.146	0.096	0.144	0.118	0.176	0.336	0.153	0.102	0.476	0.135	0.184	0.109
	(0.165)	(0.125)	(0.162)	(0.120)	(0.159)	(0.135)	(0.180)	(0.267)	(0.157)	(0.122)	(0.308)	(0.145)	(0.208)	(0.131)
μ_{4p}	0.361	0.184	0.127	0.147	0.161	0.110	0.149	0.157	0.172	0.144	0.126	0.389	0.122	0.148
	(0.279)	(0.178)	(0.143)	(0.143)	(0.161)	(0.117)	(0.171)	(0.164)	(0.175)	(0.163)	(0.139)	(0.282)	(0.141)	(0.157)
μ_{5p}	0.117	0.177	0.147	0.392	0.296	0.305	0.106	0.108	0.102	0.188	0.095	0.086	0.087	0.136
	(0.156)	(0.207)	(0.164)	(0.282)	(0.233)	(0.252)	(0.157)	(0.134)	(0.136)	(0.208)	(0.131)	(0.126)	(0.132)	(0.150)
μ_{6p}	0.090	0.135	0.092	0.183	0.087	0.185	0.081	0.169	0.095	0.156	0.083	0.124	0.381	0.411
	(0.124)	(0.169)	(0.112)	(0.195)	(0.107)	(0.182)	(0.119)	(0.197)	(0.115)	(0.151)	(0.126)	(0.150)	(0.263)	(0.301)

TABLE 5.B.5. Means of the infinite Gaussian mixtures for the German stock markets. The number in parenthesis are the standard deviation calculated from the covariance matrix Ω in each dimension. The size of each class l were n_l= 69, 36, 52, 63, 56, 57.

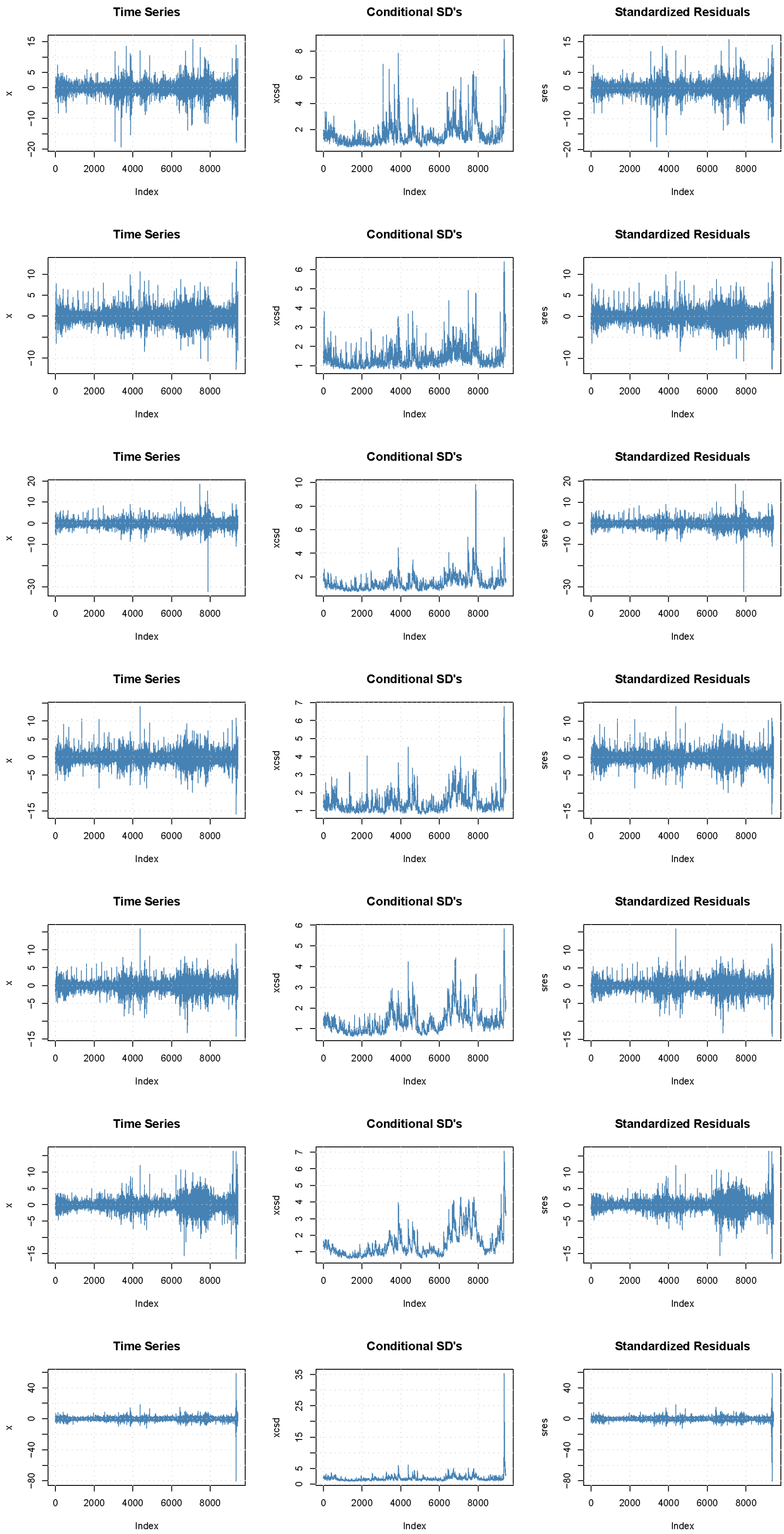

FIGURE 5.B.1. Returns (right), estimated standard deviations (middle) and the standarized residuals (left) for the stock markets under study. From top to bottom: Allianz, BASF, Bayer, E.ON, RWE, Siemens and Volkswagen Group.

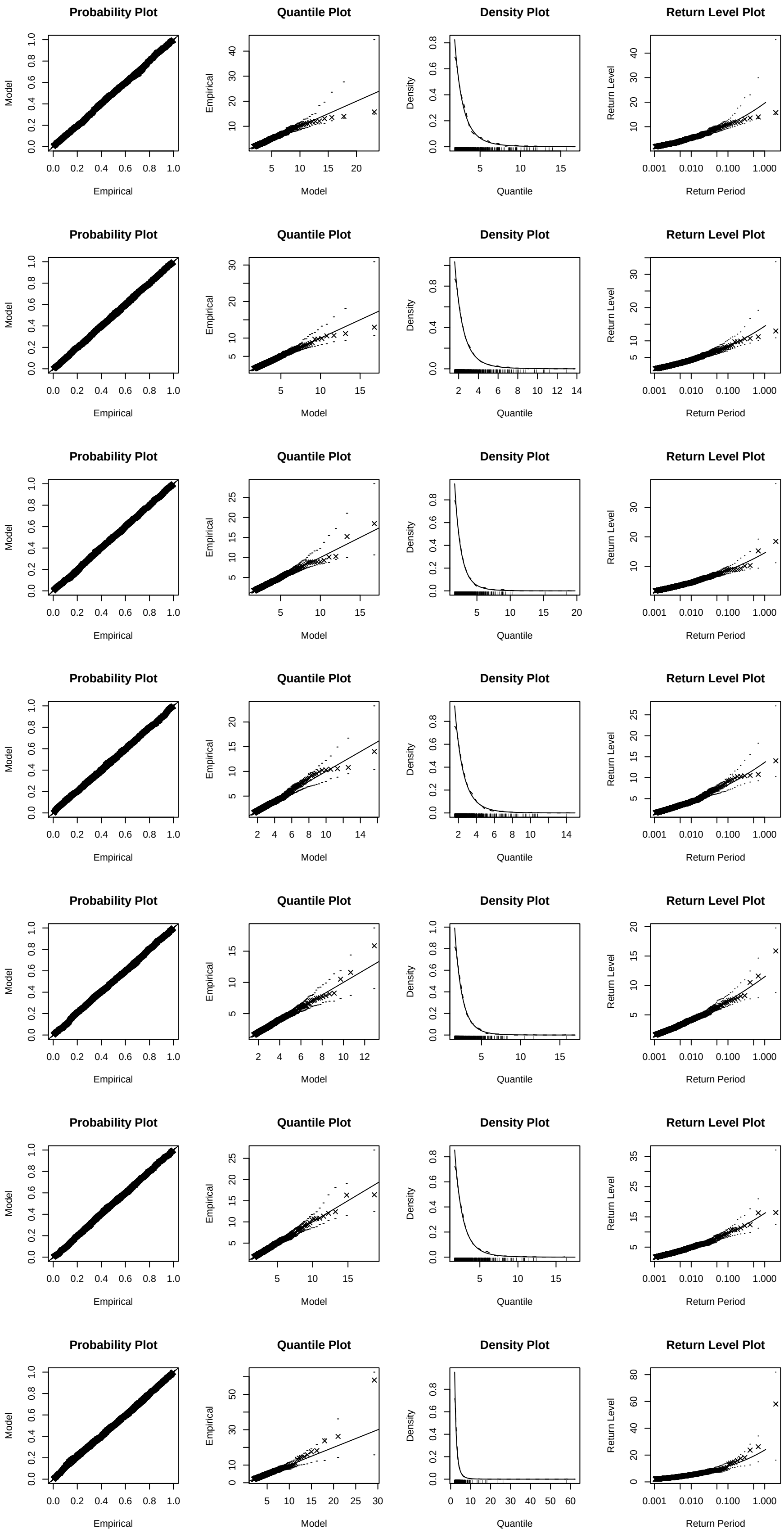

FIGURE 5.B.2. Diagnostisc plots for the GPD Models. The plots compare the parametric distribution, densities, and quantiles to their empirical counterparts. From top to bottom: Allianz, BASF, Bayer, E.ON, RWE, Siemens and Volkswagen Group.

CHAPTER 6

Topics in multivariate regular variation in finance

6.1. Introduction

Recent research in risk management has highlighted the importance of the conditional correlation as measure of contagion or interdependence, and the estimation of spillover probabilities in some financial crisis. The present chapter will offer some insight into these issues from the point of view of the theory of multidimensional extremes.

On one hand the conditional correlation lies at the heart of the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), where its use as a measure of dependence between financial instruments is essentially founded on an assumption of multivariate normally distributed returns. Increasingly, however, correlation is being used as a dependence measure in general risk management, often in areas where the assumption of multivariate normal risks is completely untenable.

On the other hand, in practical financial applications one sometimes encounters data sets with a few extremely large observations and one is concerned about the eventuality of future data points lying far out. Moreover, the influence of these few events is enormous for analyzing, for example, the risk of a portfolio. For such time series, assuming some regularity at infinity, it might be appropriate to use regular varying distributions to model underlying uncertainty. Clearly, an important difference between the multivariate and the univariate case, when analyzing extremes, is the possibility to have dependence among the components. Large values may for instance tend to occur simultaneously in different asset markets.

The conventional multivariate extreme value theory has emphasized the asymptotically dependent class resulting in its wide use in all the finance applications. However, if the series are truly asymptotically independent, such an approach will result in the over estimation of extreme value dependence and consequently of the measure of extreme risk.

In particular, this degree of asymptotic independence is directly related to the over estimation of different measures of risk. Despite this potential drawback, the case of asymptotically independent models has so far been missing from the finance literature. One first way to remedy it is by means of hidden regular variation (Heffernan and Resnick (2005); Maulik and Resnick (2004); Resnick (2006)), which measures variables on a different scale. Another is via conditioning on one component being large and using a limiting distribution as the conditioning variable is being pushed to infinity (Heffernan and Tawn (2004); Hefferman and Resnick (2007); Das and Resnick (2008)). In the first approximation we concentrate on overcoming the limitations in the case of asymptotic independence.

The contribution of this chapter is divided principally in two parts. First, we demonstrate how the conditional correlation can conduce to erroneous conclusions about extreme events. We show by means of a simple linear factor model as is done in the literature of contagion (Baig and Goldfajn (1999); Forbes (2002); Arestis et al. (2005)), that the assumption of a specific distribution function for the common risk factor directly affects the conditional correlation; hence this is not a reliable metric of dependence. We make this point more clear, through of several examples.

In particular, this chapter focuses on regular varying distributions with index $\alpha > 2$, which is equivalent to distributions whose maximum domain of attraction (MDA) of the maxima is the Frèchet distribution $\Phi_\alpha = \exp\{-x^{-\alpha}\}$ and an extension for distributions whose MDA for the maxima is in the domain of the Gumbel distribution $\Lambda = \exp\{-\exp\{-x\}\}$.

Furthermore, we provide theoretical arguments suggesting that the strong conclusion of "contagion" or "interdependence", obtained by the literature based on bivariate correlation tests, follows from arbitrary assumptions, especially on the distribution of the random variable for the common risk factor, which biases the results obtained.

Second, a semi-parametric model estimation is proposed for the asymptotic dependence and independence case in multivariate extreme value theory. This model is based on power transformations of the marginals and on a scaling property of exponent measures. The concept of asymptotic independence used in this chapter is stronger than the asymptotic independence used generally in extreme value theory, but weaker than independence.

Other contribution of this chapter is to examine the possible contagion during the Russian crisis and the more recent crisis. We apply and test our methodology in two key markets, namely stock and bond.

The empirical investigation reveals some degree of asymptotic independence among these markets, which should implicate no significant contagion effect between bond and stock exchange markets of Russia to Brazil. Nevertheless, there exist significant spillover probabilities during the Russian default, and the more recent crisis.

In particular, using daily returns on stock and bond markets from September, 1995 until January, 2001, we find tail indices to be usually stronger. In addition, with the use of a t-GARCH(1,1) filter for heterokesdasticity, we find that the tail indices are reduced considerably.

We start the chapter by reviewing the theory of regular variation in section 6.2. Section 6.3 presents some extensions to multivariate regular variation and hidden regular variation. In section 6.4 discusses the main contributions of this chapter. In particular, subsection 6.4.1 demonstrates the implications of assuming regular varying distributions in the risk factor in a linear factor model, while subsection 6.4.2 develops estimators for spillover probabilities under asymptotic independence and dependence. Data and empirical results on the Russian-Brazil contagion are present in section 6.5. Finally, section 6.6 concludes.

6.2. Regular variation

In a first approximation one may describe a regular varying distribution at infinity as one whose tails are much heavier than those of the normal or exponential distributions. Historically, a precise definition as "heavy" or " light" tails very much depends on the area of application and the structural properties of the time series one wants to model.

Experience has shown that the tails of financial data sets are typically much heavier than in the Gaussian case. Since financial crises are triggered by exceptionally large losses, much emphasis is placed nowadays on quantifying the probability of such exceptionally large losses.

Two classes of distributions have gained particular popularity for modelling extreme events: regularly varying distributions and subexponential distributions. This chapter concentrates only on the first class, for which many real-life data sets in teletraffic, insurance and finance exist empirical evidence in favour.

Nowadays, there are several books on this theory (Bingham et al. (1987); Resnick (1987); Embrechts et al. (1997); Resnick (2006)). The following is a standard definition of regular variation.

DEFINITION 6.2.1. (Regular variation) A measurable function $F:\mathbb{R}_+\to\mathbb{R}_+$ is regularly varying at ∞ with index $\alpha\in\mathbb{R}$ (we write $F\in\mathcal{R}_\alpha$) if

$$\lim_{t\to\infty}\frac{F(tx)}{F(t)}=x^{\alpha}. \tag{6.2.1}$$

If $\alpha=0$ we say that F is *slowly varying (at ∞).* Slowly varying functions are generically denoted by L. If $F\in\mathcal{R}_\alpha$, then $F(x)/x^{\alpha}$ can always be represented as $F(x)=L(x)\,x^{\alpha}$.

Let us try to give the intuition behind the definition of regular variation in terms of quantification the risk for a simple asset with distribution function $F(x)$. Equation (6.2.1) allows to replace a high threshold tx, where $t\to\infty$ and $x>0$, containing few observations by a threshold t containing many observations, and on which the distribution function has the same tail index $\alpha>0$.

An important result is the fact that convergence in (6.2.1) is *uniform*[1] on each compact subset of $(0,\infty)$.

Notice that distributional tails of type (6.2.1) are a slight generalization of distributions with pure power law tails.

EXAMPLE 6.2.2. Typical examples of slowly varying functions are positive constants or functions converging to a positive constant, logarithms and iterated logarithms. For instance, for all real α the following distributions in terms of their densities are regularly varying at ∞ with index α.

$$\begin{aligned}
t\text{-Student with } \nu \text{ d.o.f}\quad f(x) &= \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\frac{\left(1+x^2/\nu\right)^{-(\nu+1)/2}}{\sqrt{\nu\pi}},\qquad x\in\mathbb{R}\\
\text{Cauchy}\quad f(x) &= \frac{1}{\pi\left(1+x^2\right)},\quad x\in\mathbb{R}\\
\text{Pareto}\quad f(x) &= \frac{(\alpha+1)\,k^{\alpha+1}}{x^{\alpha+1}},\quad x\ge k,\quad k,\alpha>0.
\end{aligned}$$

Other important examples are stationary stochastic processes, such as the ARCH, GARCH, EGARCH, SV, etc, which have been largely used as models for financial returns (see for example Davis and Mikosch (1998, 2006a,b); Embrechts et al. (1997)).

[1] *For functions in $\mathbb{R}$, the phrase uniform convergence means that for real-valued functions $\{f_n\}_{n\ge0}$*

$$\sup_{x\in A}\left|f_0(x)-f_n(x)\right|\to 0$$

holds for any compact interval $A\subset\mathbb{R}$ as $n\to\infty$.

The most important property of regularly varying functions is the fact that the limit (6.2.1), which is assumed to exist only pointwise, holds in fact uniformly on compact sets in $\mathbb{R}$.

The next results examine the integral and differentiation properties of the regularly varying functions. The main idea in the first theorem is that integrals of regularly varying functions are again regularly varying, or more precisely, one can take the slowly varying function out of the integral. We assume all functions are locally integrable on intervals including 0 as well.

THEOREM 6.2.3. *(Karamata's Theorem)*
Suppose that $F \in \mathcal{R}_\alpha$ and F is locally bounded on $[x_0, \infty)$ for some $x_0 \geq 0$. Then,

(1) for $\alpha \geq -1$, $\int_{x_0}^{x} F(t)\, dt \in \mathcal{R}_{\alpha+1}$ and
$$\lim_{x\to\infty} \frac{\int_{x_0}^{x} F(t)\, dt}{xF(x)} = \alpha + 1$$
(2) for $\alpha < -1$ (or if $\alpha = -1$ and $\int_{x}^{\infty} F(s)\, ds < \infty$), then $F \in \mathcal{R}_\alpha$ implies that $\int_{x_0}^{x} F(t)\, dt \in \mathcal{R}_{\alpha+1}$ and
$$\lim_{x\to\infty} \frac{\int_{x_0}^{x} F(t)\, dt}{xF(x)} = -(\alpha + 1)$$

The following result is crucial for the differentiation of regularly varying functions.

THEOREM 6.2.4. *(Monotone density Theorem)*
Let $F(x) = \int_0^x f(s)\, ds$ with f monotone on $(z, \infty)$for some $z > 0$. If
$$F(x) \sim cx^\alpha L(x), \quad x \to \infty, \tag{6.2.2}$$
with $c \geq 0$, $\alpha \in \mathbb{R}$ and L is slowly varying, then
$$f(x) \sim c\alpha x^{\alpha-1} L(x), \quad x \to \infty. \tag{6.2.3}$$

For $c = 0$ the above relations are interpreted as $F(x) \sim= o(x^\alpha L(x))$ and $f(x) = o\left(x^{\alpha-1} L(x)\right)$.

The applicability of regular variation is further enhanced by the Karamata representation and it will be of very much help in subsection 6.4.1 for the derivation of conditional correlations.

COROLLARY 6.2.5. *(the Karamata's representation)*

(1) *The function L is slowly varying, if and only if (iff), L admits the representation*
$$L(x) = c(x) \exp\left\{\int_1^x t^{-1} \varepsilon(t)\, dt\right\},\ x > 0, \tag{6.2.4}$$
where $c : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $\varepsilon : \mathbb{R}_+ \mapsto \mathbb{R}_+$, and
$$\lim_{x\to\infty} c(x) = c \in (0, \infty), \tag{6.2.5}$$
$$\lim_{t\to\infty} c(t) = 0.$$

(2) *A function* $F : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *is regularly varying with index* α *iff* F *has representation*

$$F(x) = c(x) \exp\left\{ \int_1^x t^{-1} \alpha(t)\, dt \right\}, \ x > 0,$$

where $c(\cdot)$ *satisfies (6.2.5 on the preceding page) and* $\lim_{t\to\infty} \alpha(t) = \alpha$.

Other interesting form of regular variation is Rapid variation

Definition 6.2.6. (Rapid variation)

We say a Lebesgue measurable function h in $(0, \infty)$ is rapidly varying with index $-\infty$, i.e., $h \in \mathcal{R}_{-\infty}$ if

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = \begin{cases} 0 & if \quad t > 1, \\ \infty & if \quad 0 < t < 1. \end{cases}$$

An example of a function $h \in \mathcal{R}_{-\infty}$ is $h(x) = \exp(-x)$.

The next proposition summarizes some of the results which are useful for tails of distribution functions.

Proposition 6.2.7. *(Regular variation for tails of distribution functions (Embrechts et al. (1997)))*

Suppose F *is a distribution function with* $F(x) < 1$ *for all* $x \geq 0$.

(1) If the sequences $\{a_n\}$ and $\{x_n\}$ satisfy $a_n/a_{n+1} \to 1$, $x_n \to \infty$, and if for some real function g and all λ form a dense subset of $(0, \infty)$,

$$\lim_{n\to\infty} a_n \overline{F}(\lambda x_n) = g(\lambda) \in (0, \infty),$$

then $g(\lambda) = \lambda^{-\alpha}$ for some $\alpha \geq 0$ and $\overline{F}(x)$ is regularly varying.

(2) Suppose F is absolutely continuous with density f such that for some $a > 0$, $\lim_{x\to\infty} xf(x)/\overline{F}(x) = \alpha$. Then $f \in \mathcal{R}_{-(1+\alpha)}$ and consequently $\overline{F} \in \mathcal{R}_{-\alpha}$.

(3) Suppose $f \in \mathcal{R}_{-(1+\alpha)}$ for some $\alpha > 0$. Then, $\lim_{x\to\infty} xf(x)/\overline{F}(x) = \alpha$.

(4) Suppose X is non-negative random variable with distribution tail $\overline{F} \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$. Then, $EX^\beta < \infty$ if $\beta < \alpha$, and $EX^\beta = \infty$ if $\beta > \alpha$.

(5) Suppose $\overline{F} \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$, $\beta \geq \alpha$. Then

$$\lim_{x\to\infty} \frac{\int_0^x t^\beta dF(t)}{x^\beta \overline{F}(x)} = \frac{\alpha}{\beta - \alpha}.$$

The converse also holds in the case that $\beta > \alpha$. If $\beta = \alpha$ one can only conclude that $\overline{F}(x) = o(x^{-\alpha} L(x))$ for some $L \in \mathcal{R}_0$.

When one is working with convergence of probability measures in heavy tailed analysis a modern theory of weak convergence of probability measures on metric spaces is necessary to clearly understand asymptotic properties of some statistics. The vague topology is an example of the weak topology which arises in the study of measures on locally compact Hausdorff spaces (see Resnick (2006)).

Definition 6.2.8. (*Vague Convergence*) Define the space $(\mathbb{C}, \mathcal{E})$, which is locally compact with countable base and $M_+(\mathbb{C})$ is the space of all Radon measures[2] on $\mathbb{C}$. So $\mu \in M_+(\mathbb{C})$

[2]Radon measure is defined in measure theory to be a measure on the ς-algebra of some Borel set that is locally finite and inner regular.

means that μ is a measure on $\mathcal{E}$ and $\mu(K) < \infty$ for all compact sets $K \in \mathcal{E}$. Denote $C_K^+(\mathbb{C}) := \{f : f : \mathbb{C} \to \mathbb{R}_+, f \text{ continuos with compact support}\}$. Then, if $\mu_n, \mu \in M_+(\mathbb{C})$, $\mu_n \xrightarrow{v} \mu$ if for every $f \in C_K^+(\mathbb{C})$,

$$\mu_n(f) := \int_{\mathbb{C}} f d\mu_n \to \int_{\mathbb{C}} f d\mu =: \mu(f).$$

The following theorem resumes regular variation results of one dimensional distribution function tails. For a proof, we refer to Resnick (2006). Note the use of the cone $\mathbb{C} = (0, \infty]$ in the next theorem, which is a *one point uncompactification* of the compact set $[0, \infty]$ (see Resnick (2006), pp. 170). The idea behind of this topology is to make neighborhoods of ∞ relatively compact. This is necessary because vague convergence only controls behavior of measures on relatively compact sets and $(0, \infty]$ is a natural set when dealing with right tail problems. Note that we are working in the positive orthant. However, one can get the same results in other directions.

THEOREM 6.2.9. *(Regular Variation in one dimensional distribution function tails). Let X be a random vector with distribution function F and tail $\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ and quantile function $b_n := b(n) = \left(\bar{F}(x)^{-1}\right)^{\leftarrow}(n) = F^{\leftarrow}\left(1 - n^{-1}\right)$. Then, the following results are equivalent:*

(1) $\bar{F} \in \mathcal{R}_{-\alpha}$, $\alpha > 0$.
(2) *There exists $b_n \to \infty$ such that $n\bar{F}(b_n x) \to x^{-\alpha}$, $x > 0$.*
(3) *There exists $b_n \to \infty$ such that $n\mathbb{P}\left(b_n^{-1} X \in \cdot\right) \xrightarrow{v} v_\alpha(\cdot)$, in $M_+((0, \infty])$ satisfies $v_\alpha(x, \infty] = x^{-\alpha}$, for $x > 0$.*

EXAMPLE 6.2.10. Consider a random vector X with distribution function $\overline{F} \in \mathcal{R}_\alpha$ on the cone $\mathbb{C}$. Then, by Theorem 6.2.9 we have $n\mathbb{P}\left(b_n^{-1} X > x\right) \xrightarrow{v} v_\alpha((x, \infty])$ for $x \to \infty$, and the sequential form of regular variation yields to $v_\alpha(x, \infty] = c_+ x^{-\alpha}$ with $\alpha > 0$ and $c_+ \geq 0$.

Similarly, observe that $n\mathbb{P}\left(b_n^{-1} X \leq -x\right) \xrightarrow{v} v_\alpha([-\infty, -x])$, which implies $v_\alpha((x, \infty]) = c_- x^{-\alpha}$ for $\alpha > 0$ and $c_- \geq 0$. Note that the only issue here are the constants c_+ and c_-, by the fact that index of regular variation α is the same for both tails. For this it is enough to observe that $b_n \in \mathcal{R}_{1/\alpha}$.

Further, the limit $n\mathbb{P}\left(b_n^{-1} |X| > x\right) \in \mathcal{R}_{-\alpha}$ yields to

$$\lim_{t \to \infty} \frac{\mathbb{P}(X > t)}{\mathbb{P}(|X| > t)} = \frac{c_+}{c_+ + c_-} =: p, \text{ and } \lim_{t \to \infty} \frac{\mathbb{P}(X < -t)}{\mathbb{P}(|X| > t)} = \frac{c_-}{c_+ + c_-} =: q \tag{6.2.6}$$

and $p + q = 1$.

The result in equation (6.2.6) is referred as *tail balance condition*.

Note that all components of X are regularly varying with the same tail index α, which is guaranteed by the fact that $b_n \in \mathcal{R}_{1/\alpha}$. This is the standard case of multivariate regular variation, though we do not rule out the possibility that one or more of the coefficients c can be zero. This occurs if the components of X are regularly varying with different tail indices, a natural scenario in practice. In this case the spectral measure will place mass only along the axis or axes of lowest α, as is illustrated in the next example.

EXAMPLE 6.2.11. Let $\mathbf{X} = \{X_1, X_2\}$ be a bivariate random vector with tail indices α and $\alpha + \Delta$, where $\triangle > 0$. In particular we know that $\overline{F}_1(x_n) \sim x_n^{-\alpha} \mathcal{L}_1(x_n)$ and $\overline{F}_2(x_n) \sim$

$x_n^{-(\alpha+\triangle)}\mathcal{L}_2(x_n)$, where $\mathcal{L}_1, \mathcal{L}_2$ are slowly varying functions. Then

$$\begin{aligned} n\overline{F}(x_n, x_n) &= n\mathbb{P}(\|\mathbf{X}\|_{\max} > x_n) \\ &= n\{1 - \mathbb{P}(X_1 \vee X_2 \leq x_n)\} \\ &= nx_n^{-\alpha}\mathcal{L}_1(x_n)\left\{1 + \frac{x_n^{-\Delta}\mathcal{L}_2(x_n)}{\mathcal{L}_1(x_n)(\mathcal{L}_1(x_n) - 1)}\right\} \\ &\sim nx_n^{-\alpha}\mathcal{L}_1(x_n). \end{aligned}$$

Thus, the spectral measure for different tail indices is reduced to a single point mass along the axis with the smallest tail index.

6.3. Multivariate regular variation in cones

In practice one is often confronted with multivariate problems on a variety of stock or bond indices, or, perhaps the same assets in different markets. A multivariate concept of regular variation then will be needed. However, the extension of a mathematical notion from the one dimensional to the higher dimensional case often leads to a great variety of different approximations. The great majority of the results to be presented is known and can be found in Bingham et al. (1987); Resnick (1987); Embrechts et al. (1997); Davis and Mikosch (1998); Maulik and Resnick (2004); Heffernan and Resnick (2005). The intention in this section is to present a unificated approach to the multivariate regular variation of random vectors. Moreover, in this section operations between vectors should be interpreted componentwise.

Definition 6.3.1. A subset $\mathbb{C} \subset \mathbb{R}^d$ is a cone if whenever $\mathbf{x} \in \mathbb{C}$ also $t\mathbf{x} \in \mathbb{C}$ for any $t > 0$. Furthermore, a function $h : C \to \mathbb{R}_+$ is monotone if it is either non-decreasing in each component or non-increasing in each component. For h non-decreasing is equivalent to saying that whenever $\mathbf{x}, \mathbf{y} \in \mathbb{C}$ and $\mathbf{x} \leq \mathbf{y}$ we have $h(\mathbf{x}) \leq h(\mathbf{y})$. The natural domain for a multivariate regularly varying function is a cone.

A function $h : \mathbb{C} \to \mathbb{R}_+$ is regularly varying at ∞ with limit function $\lambda(\mathbf{x}) > 0$, if for all $\mathbf{x} \in \mathbb{C}$,

$$\lim_{t\to\infty} \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \lambda(\mathbf{x}) > 0, \quad \mathbf{x} \in \mathbb{C}.$$

An inmediate implication is that the limit function $\lambda(\mathbf{x})$ is homogeneous

$$\lambda(s\mathbf{x}) = s^{\rho}\lambda(\mathbf{x}), \ s > 0, \mathbf{x} \in \mathbb{C}, \ \rho \in \mathbb{R}.$$

For this result it is enough to observe for a function $U(t) = h(t\mathbf{x})$ for $t > 0$ we have that $U \in \mathcal{R}_\rho$ for some $\rho \in \mathbb{R}$

$$\lim_{t\to\infty} \frac{U(ts)}{U(t)} = \lim_{t\to\infty} \frac{h(ts\mathbf{x})}{h(t\mathbf{x})} = \lim_{t\to\infty} \frac{h(ts\mathbf{x})\,h(t\mathbf{1})}{h(t\mathbf{1})\,h(t\mathbf{x})} = \frac{\lambda(s\mathbf{x})}{\lambda(\mathbf{x})} = s^{\rho}.$$

6.3.1. Multivariate regular variation of tail probabilities. In a slightly different terminology, the following characterization of multivariate regular variation is done in the cone $\mathbb{C} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$. This space is the *one-point uncompactification of* $[\mathbf{0}, \infty]$. The exclusion of $\{\mathbf{0}\}$ is necessary by the fact that we study regular variation at infinity, i.e., compact subsets bound away from $\mathbf{0}$.

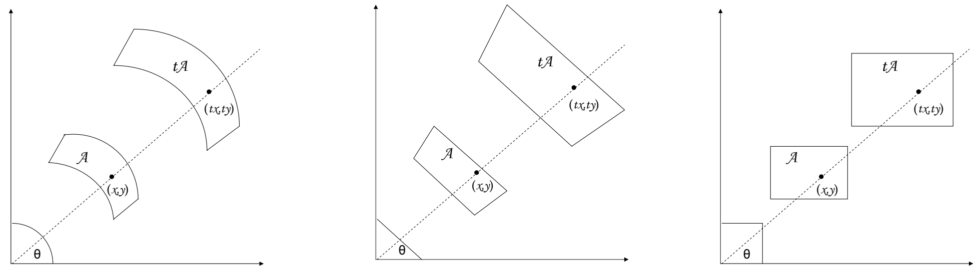

FIGURE 6.3.1. Different types of metrics in $\mathbb{R}_+^2$: (left) Euclidean norm, (middle) Sum-Morm, (right) Max-Norm.

DEFINITION 6.3.2. Let $\mathbf{X}$ be a nonnegative random vector with the regular varying distribution function $F \in \mathcal{R}_\alpha$ concentrate on $(\mathbf{0}, \infty]$. The following formulations of regular variation are equivalent

(1) *(Global and marginal condition)* There exist $\mathbf{b}(t) \to \infty$ and a Radon measure v on $\mathbb{C}$ such that in $M_+(\mathbb{C})$, the space of positive Radon measures on $\mathbb{C}$.

$$t\mathbb{P}\left(\mathbf{b}(t)^{-1}\mathbf{X} \in \cdot\right) \xrightarrow{v} v_\alpha(\cdot)\ ,\ t \to \infty. \tag{6.3.1}$$

Then, $b_j(t) \in \mathcal{R}_{1/\alpha_j}$ the class of regularly varying functions with $\alpha_j > 0$, for all $j = 1, \ldots, d$. Furthermore, the marginal convergences satisfy

$$t\mathbb{P}\left(b(t)^{-1} X_j > x\right) \to v_{\alpha_j}(x, \infty] := x^{-\alpha_j}. \tag{6.3.2}$$

(2) *(Spectral measure and polar transformation)* Let $\|\cdot\|$ be a norm in $\mathbb{R}^d$ and denote the unit sphere in this norm by $\mathbb{S}^{d-1} := \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\right\}$. Furthermore, let $\mathbb{S}_+^{d-1} = \mathbb{S}^{d-1} \cap [\mathbf{0}, \infty]^d$. Then, there exists a probability measure $\mathcal{S}(\cdot)$ on $\mathbb{S}_+^{d-1}$ and a sequence $b(t) \to \infty$ such that for $(R, \Theta) = \left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|}\right)$ we have

$$t\mathbb{P}\left(\left(b(t)^{-1} R,\ \Theta\right) \in \cdot\right) \xrightarrow{v} cv_\alpha(\cdot) \times \mathcal{S} \tag{6.3.3}$$

in $M_+\left((0, \infty] \times \mathbb{S}_+^{d-1}\right)$.

(3) *(The Poisson transform)* There exists $b_n \to \infty$ such that

$$\sum_{i=1}^{n} \epsilon_{b_n^{-1}\mathbf{X}} \Rightarrow PRM(v) \tag{6.3.4}$$

and

$$\sum_{i=1}^{n} \epsilon_{\left(b_n^{-1}R, \Theta\right)} \Rightarrow PRM(cv_\alpha \times \mathcal{S}) \tag{6.3.5}$$

in $M_p(\mathbb{C})$ and $M_p\left((0, \infty] \times \mathbb{S}_+^{d-1}\right)$ respectively.

(4) *(Convergence of empirical measures)* The above implies that for any sequence $k \to \infty$ such that $n/k \to \infty$ one obtains

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{b(n/k)^{-1}\mathbf{X}} \Rightarrow v_\alpha \quad \text{in } M_+(\mathbb{C}) \tag{6.3.6}$$

and

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(b_n^{-1}R,\Theta\right)} \Rightarrow cv_\alpha \times \mathcal{S} \quad \text{in } M_+\left((0,\infty] \times \mathbb{S}_+^{d-1}\right). \tag{6.3.7}$$

Intuitively, regular variation means that asymptotically the distribution in polar coordinates can be represented by a product measure, which is the product of the spectral measure and a radial measure, which has power decay.

The characterization of multivariate regular variation in (6.3.1) is referred to as the sequential definition of regular variation (see Embrechts et al. (1997)) and it describes dependence among the marginals, while (6.3.2) rules out tails that are not regular varying. It is interesting to note that whereas univariate regular variation has been introduced as a function property, all the characterizations of multivariate regular variation are in terms of measures. Multivariate distribution functions are de facto more difficult to deal with than univariate distribution functions and therefore measures appear more natural.

The relation (6.3.3) in terms of coordinates polar allows to estimate the spectral measure $\mathcal{S}$, which describes the way how extreme movements of univariate marginals are related to each other. Knowledge of this measure facilities the estimation of joint and conditional probabilities.

The relations (6.3.4) and (6.3.5) show as multivariate regular variation is equivalent to induced empirical measures weakly converging to Poisson random measure limits (for a proof of this statemen see Resnick (2006, page 180) or Embrechts et al. (1997, page 232)). The advantage of this method is the dimensionless aspect. Thus, this method works just as well in $\mathbb{R}^d$ as in $\mathbb{R}$.

Relations (6.3.6) and (6.3.7) show consistent estimators of the measure $v_\alpha \in M_+(\mathbb{C})$, provided $n \to \infty$, $k/n \to \infty$, and as soon as a consistent estimator of $b(k/n)$ is specified.

Finally, we do not defined which norm $\|\cdot\|$ in $\mathbb{R}^d$ to use. This is because of the fact that all norms in $\mathbb{R}^d$ are equivalent. An immediate consequence of Definition 6.3.2 is then the following result.

COROLLARY 6.3.3. *Let $\|\cdot\|_A$ and $\|\cdot\|_B$ be two norms on $\mathbb{R}^d$ and let $\mathbf{X}$ be a d-variate random vector. Then $\mathbf{X}$ is regularly varying with index α in respect to the norm $\|\cdot\|_A$ if and only if $\mathbf{X}$ is regularly varying with index with respect to the norm$\|\cdot\|_B$.*

However, it is clear that the spectral measures will not coincide for different norms. The next example for bivariate elliptical distributions illustrates this idea in respect to the max-norm $\|\cdot\|_{\max}$ and the Euclidean norm $\|\cdot\|_2$. In practical applications the choice of norm should be related to the questions we are trying to answer. As for example in the financial stock markets the question can be: Given that at least one of the two stock markets have experienced a crash, which is the probability of an extreme movement in the other stock market index? In this case the max-norm should be used.

EXAMPLE 6.3.4. Let $\mathbf{X}$ be a regularly varying bivariate elliptical distribution with index α and stochastic representation $\mathbf{X} \stackrel{d}{=} RAU$ or matrix format

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \stackrel{d}{=} R \begin{pmatrix} \sqrt{\Sigma_{11}} & 0 \\ \sqrt{\Sigma_{22}}\rho_{12} & \sqrt{\Sigma_{22}\left(1-\rho_{12}^2\right)} \end{pmatrix} \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix},$$

where $\varphi \sim U\left(-\frac{\pi}{2}, \frac{3\pi}{2}\right)$. Let

$$f(x) = \sqrt{\Sigma_{11}\cos^2 t + \Sigma_{22}\sin^2(\arcsin\rho_{12}+\varphi)},$$

$$g(t) = \begin{cases} -\pi/2 & t = -\pi/2, \\ \arctan\left(\sqrt{\frac{\Sigma_{22}}{\Sigma_{11}}}\left(\rho_{12} + \sqrt{1-\rho_{12}^2}\tan t\right)\right) & t \in (-\pi/2, \pi/2), \\ g(t-\pi)+\pi & t \in [\pi/2, 3\pi/2). \end{cases}$$

Since R is regularly varying which implies that $\mathbb{P}(R > x) = x^{-\alpha}L(x)$, where $L(x)$ is a slowly varying function, we can obtain the spectral measure $S_{\theta_1\theta_2}\left\{(\cos t, \sin t)^t : \theta_1 < t < \theta_2\right\}$ with euclidean norm as

$$\begin{aligned} \lim_{z\to\infty} \frac{\mathbb{P}\left(R\|AU\|_2 > zx,\ AU/\|AU\|_2 \in S\right)}{\mathbb{P}\left(R\|AU\|_2 > z\right)} &= \lim_{z\to\infty} \frac{\int_{g^{\leftarrow}(\theta_1)}^{g^{\leftarrow}(\theta_2)} (zx)^{-\alpha} f(x)^{\alpha} L(zx/f(t))\,dt}{\int_0^{2\pi} z^{-\alpha} f(x)^{\alpha} L(z/f(t))\,dt} \\ &= x^{-\alpha} \lim_{z\to\infty} \frac{\int_{g^{\leftarrow}(\theta_1)}^{g^{\leftarrow}(\theta_2)} f(x)^{\alpha} L(zx/f(t))\,dt}{\int_0^{2\pi} f(x)^{\alpha} L(z/f(t))\,dt} \\ &= x^{-\alpha} \lim_{z\to\infty} \frac{\int_{g^{\leftarrow}(\theta_1)}^{g^{\leftarrow}(\theta_2)} f(x)^{\alpha}\,dt}{\int_0^{2\pi} f(x)^{\alpha}\,dt}, \end{aligned}$$

where the third equality follows from the fact that $L(tx)/L(t) \to 1$. This yields to the spectral measure

$$\mathcal{S}([\theta_1, \theta_2]) = \frac{\int_{g^{\leftarrow}(\theta_1)}^{g^{\leftarrow}(\theta_2)} \left\{\Sigma_{11}\cos^2 t + \Sigma_{22}\sin(\arcsin\rho_{12}+t)\right\}^{\alpha/2} dt}{\int_0^{2\pi} \left\{\Sigma_{11}\cos^2 t + \Sigma_{22}\sin(\arcsin\rho_{12}+t)\right\}^{\alpha/2} dt}.$$

Proceeding analogously for the spectral measure with respect to the max-norm $\|\cdot\|_{\max}$ but with function

$$f(x) := \max\left\{\sqrt{\Sigma_{11}}\,|\cos t|, \sqrt{\Sigma_{22}}\,|\sin(\arcsin\rho_{12}+t)|\right\}$$

we find that

$$\mathcal{S}([\theta_1, \theta_2]) = \frac{\int_{g^{\leftarrow}(\theta_1)}^{g^{\leftarrow}(\theta_2)} \max\left\{\sqrt{\Sigma_{11}}\,|\cos t|, \sqrt{\Sigma_{22}}\,|\sin(\arcsin\rho_{12}+t)|\right\}^{\alpha} dt}{\int_0^{2\pi} \max\left\{\sqrt{\Sigma_{11}}\,|\cos t|, \sqrt{\Sigma_{22}}\,|\sin(\arcsin\rho_{12}+t)|\right\}^{\alpha} dt}.$$

Note that these two spectral measures are absolutely continuous and hence these have density. The density for these two examples (two norms) are plotted in Figure 6.3.2 for bivariate regularly varying elliptical distributions with $(\Sigma_{11}, \Sigma_{22}) = (1, 1)$, $\rho_{12} = 0.3, 0.5, 0.8$ and with tail indices $\alpha = 1, 2, 4, 8, 16$. Notice, that as the tail become lighter the spectral measure becomes less concentrates on the diagonals of the first and third quadrant, i.e., the angels $\pi/4$ and $5\pi/4$. As consequence the probability of spillovers (for example two crashes in stock markets) become very small compared to the probability that one component is

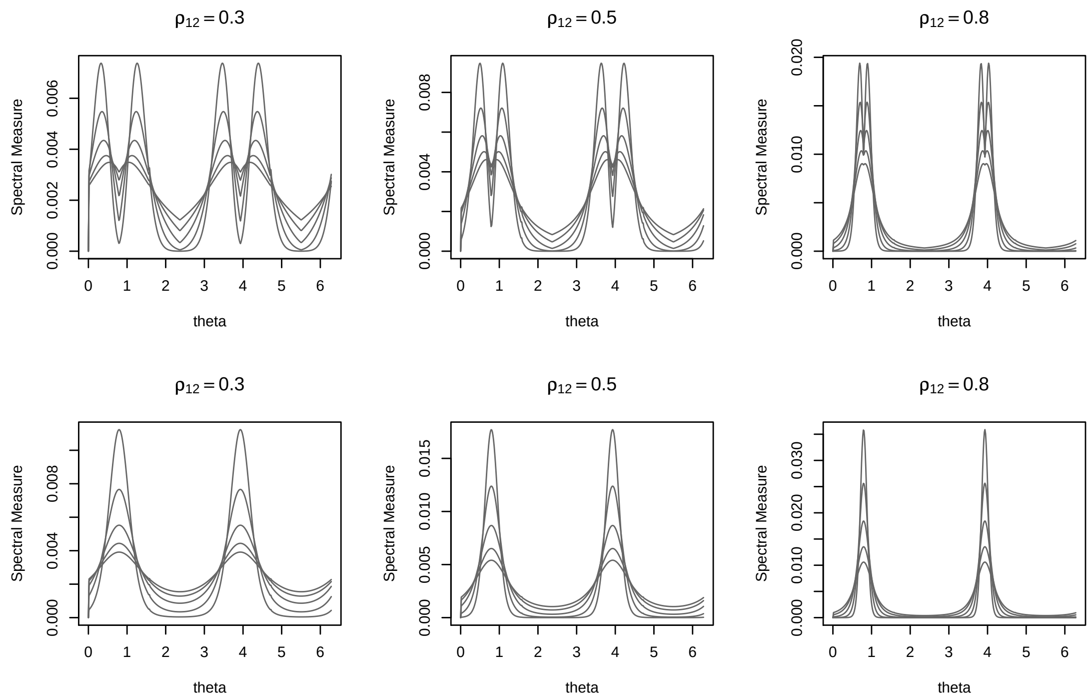

FIGURE 6.3.2. Densities of the Spectral measure of Example 6.3.4 with $(\Sigma_{11}, \Sigma_{22}) = (1, 1)$, $\rho_{12} = 0.3, 0.5, 0.8$ and with tail indices $\alpha = 1, 2, 4, 8, 16$, in respect to the max-norm (top) and the Euclidean-norm (bottom). Larger tail indices correspond to higher peaks.

extreme. In terms of tail dependence: the higher is the tail index, the smaller will become the coefficient of tail dependence.

Furthermore, the difference of the spectral measure in relation to different norms is evident, which motivated this example.

In practical applications it is clear that we cannot hope that all the marginals have the same tail index $\alpha_j > 0$. Therefore, a proper normalization is essential for handling asymptotic behaviour and concentrate on the main issues. With a change of variables regular variation with unequal components can be *standardized* for which the marginal distribution of each marginal is tail equivalent.

DEFINITION 6.3.5. (Regular variation in standard form) Let $\mathbf{X}$ be a nonnegative random vector with regular varying distribution function $F \in \mathcal{R}_\alpha$ concentrating on $(\mathbf{0}, \infty]$. The Definition 6.3.2 is equivalent in *standard form* to

(1) There exist $\mathbf{b}(t) \to \infty$ and a Radon measure μ on $\mathbb{C}$ such that in $M_+(\mathbb{C})$

$$t\mathbb{P}\left(t^{-1}\mathbf{b}^{\leftarrow}(\mathbf{X}) \in \cdot\right) \xrightarrow{v} \mu(\cdot) \ , \ t \to \infty. \tag{6.3.8}$$

Then, $b_j(t) \in \mathcal{R}_1$ the class of regularly varying functions with $\alpha_j = 1$, for all $j = 1, \ldots, d$. Furthermore, the marginal convergences satisfy

$$t\mathbb{P}\left(t^{-1}b_j^{\leftarrow}(X_j) > x\right) \to \mu\left(x^{-\alpha}, \infty\right] := x^{-1}, \tag{6.3.9}$$

where μ satisfies the homogeneity condition

$$\mu(t\cdot) = t^{-1}\mu(\cdot)$$

on Borel subsets of $\mathbb{C}$ and the measures μ and v are related by

$$v_\alpha([\mathbf{0},\mathbf{x}]^c) = \mu([\mathbf{0},\mathbf{x}^\alpha]^c), \quad x > 0.$$

(2) Let $\|\cdot\|$ be a norm in $\mathbb{R}^d$ and denote the unit sphere in this norm by $\mathbb{S}^{d-1} := \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\right\}$. Furthermore, let $\mathbb{S}_+^{d-1} = \mathbb{S}^{d-1} \cap [\mathbf{0},\infty]^d$. Then there exists a probability measure $\mathcal{S}(\cdot)$ on $\mathbb{S}_+^{d-1}$ and a sequence $b(t) \to \infty$ such that for $(R,\Theta) = \left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|}\right)$ we have

$$t\mathbb{P}\left(\left(b(t)^{-1}R,\ \Theta\right) \in \cdot\right) \xrightarrow{v} c\mu(\cdot) \times \mathcal{S}$$

in $M_+\left((0,\infty] \times \mathbb{S}_+^{d-1}\right)$.

As we said in the beginning of this section multivariate regular variation on the cone $\mathbb{C}$ forms the base of the classical extreme value theory as the next example shows.

Example 6.3.6. Let $\mathbf{X}$ be a multivariate random vector with regular varying distribution F on $\mathbb{C}$. Further, define a new vector $\mathbf{X}^* = b^{\leftarrow}(\mathbf{X})$ in *standard form* with distribution function F^*, which means that $b(t) = t$ and X^* has a limit Radon measure

$$\mu(\cdot) = t\mathbb{P}\left(X^* t^{-1} \in \cdot\right) \tag{6.3.10}$$

in $M_+([\mathbf{0},\infty] \setminus \{\mathbf{0}\})$.

Suppose $a(\cdot) > 0$ and $b(\cdot) \in \mathbb{R}$ are measure ables functions in $\mathbb{R}^+$ satisfying,

$$\frac{b(tx) - b(t)}{a(t)} \to \psi(x), \quad x > 0,\ t \to \infty, \tag{6.3.11}$$

which has the form

$$\psi(x) = \begin{cases} k\left(\frac{x^\rho - 1}{\rho}\right) & \rho \in \mathbb{R},\ \rho \neq \mathbb{R},\ x > 0 \\ k\log x & \rho = 0,\ x > 0 \end{cases}$$

for some $k \neq 0$ and where $\psi \neq 0$.

Suppose for convenience that $k = 1$, the inverse function of equation (6.3.11) yields

$$\psi^{\leftarrow}(y) \to \frac{b^{\leftarrow}(b(t) - ya(t))}{t} = (1+\rho y)^{1/\rho}, \quad 1 + \rho y > 0.$$

Let $\mathbf{a}(\cdot) > 0$ and $\mathbf{b}(\cdot) \in \mathbb{R}^d$ be a vector satisfying equation (6.3.11) for each component in $\mathbf{X}^*$, then

$$\begin{aligned} t\mathbb{P}\left(\frac{\mathbf{b}(\mathbf{X}^*) - \mathbf{b}(t)}{\mathbf{a}(t)} \nleq \mathbf{x}\right) &= t\mathbb{P}\left(\mathbf{X}^* t^{-1} \nleq \frac{\mathbf{b}^{\leftarrow}(\mathbf{b}(t) + \mathbf{x}\mathbf{a}(t))}{t}\right) \\ &\to \mu\left(\mathbf{y} \nleq \psi^{\leftarrow}(\mathbf{x})\right). \end{aligned}$$

Setting $\mathbf{X} = \mathbf{b}\mathbf{X}$ and replacing t with n

$$n\mathbb{P}(\mathbf{X} \nleq \mathbf{x}\mathbf{a}_n + \mathbf{b}_n) \quad \sim \quad n\left(-\log \mathbb{P}(\mathbf{X} \leq \mathbf{x}\mathbf{a}_n + \mathbf{b}_n)\right)$$

because of the identity $x \sim -\log(1-x)$ for $x \to 0$. Exponentianting the last equation we obtain

$$\mathbb{P}(\mathbf{X} \leq \mathbf{x}\mathbf{a}_n + \mathbf{b}_n)^n \to \exp\{-\mu(\mathbf{y} \nleq \psi^{\leftarrow}(\mathbf{x}))\}.$$

Further, if $\mathbf{X}$ are iid random vectors we conclude as $n \to \infty$ that for $\mathbf{M}_n = \bigvee_{i=1}^{n} \frac{\mathbf{X}_i - \mathbf{b}_i(t)}{\mathbf{a}_i(t)}$

$$\mathbb{P}(\mathbf{M}_n \leq \mathbf{x}) \to \exp\{-\mu(\mathbf{y} \nleq \psi^{\leftarrow}(\mathbf{x}))\}$$

the limit form of the general max-stable distributions.

6.3.2. Multivariate hidden regular variation on the cone. In the classical setting of bivariate extreme value theory, the procedures to estimate the probability of an extreme event are not applicable if the componentwise maxima of the observations are asymptotically independent. In this case, the spectral measure S will concentrate on the axes, or equivalently, the normalization will not be able to distinguish between independence or *asymptotic independence* as we will show in the next example.

EXAMPLE 6.3.7. Suppose N_1, N_2 are iid $N(0,1)$ random variables and correlation $|\rho| \leq 1$. Define $\mathbf{X} = (X_1, X_2) = \left(\sqrt{1-\rho^2}N_1 + \rho N_2, N_2\right)$ which is a bivariate normal vector $F(\cdot)$ with means 0, variances 1 and correlation ρ. Then $\lim_{t\to\infty} \mathbb{P}(X_1 > t,\ X_2 > t) = 0$, which is equivalent to asymptotic independence or that the measure $\mu(\mathbb{C}) = 0$ in $M_+(\mathbb{C})$.

Notice that

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{P}(X_1 > t,\ X_2 > t) &= \lim_{t\to\infty} \mathbb{P}\left(\sqrt{1-\rho^2}N_1 + \rho N_2 > t,\ N_2 > t\right) \\
&\leq \lim_{t\to\infty} \mathbb{P}\left(N_2 + \rho N_2 + \sqrt{1-\rho^2}N_1 > 2t\right) \\
&= \lim_{t\to\infty} \mathbb{P}\left((1+\rho) N_2 + \sqrt{1-\rho^2}N_1 > 2t\right)
\end{aligned}$$

and because $(1+\rho)N_2 + \sqrt{1-\rho^2}N_1$ is $N(0, 2(1+\rho))$ this probability is $\overline{N}\left(\frac{2t}{\sqrt{2(1+\rho)}}\right)$. Since the normal distribution belongs to the maximum domain of atraction of the Gumbel distribution we have that for all $A > 1$, $\overline{N}(At)/\overline{N}(t) \to 0$, then necessary the result follows if $A = \frac{2}{\sqrt{2(1+\rho)}} > 1$, which is the same as $2 > \sqrt{2(1+\rho)}$ and it holds for all $|\rho| \leq 1$.

Thus, the usual form to investigated multivariate dependence in extreme regions, as it is discussed in the traditional context of extreme value theory, does not apply to the case where there exist asymptotically independence between the vectors. In other words, where the probability of more than one extreme value in two components is practically zero, hence with uninformative probabilites. Furthermore, many estimators may behave badly under asymptotic independence. Indeed, the estimators based on the assumption of asymptotic dependence may be asymptotically normal with an asymptotic variance zero in this case. Thus, a refinement to the theory is necessary in the case of asymptotic independence.

This section introduces a new concept of asymptotic independence based on hidden regular variation (Maulik and Resnick (2004); Heffernan and Resnick (2005)). Hidden regular variation is a semi-parametric subfamily of the full family of distributions possessing multivariate regular variation and asymptotic independence.

Define the cones $\mathbb{C} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$ and $\mathbb{C}_0 = \{\mathbf{x} \in C : \exists\, 1 \leq i \leq j \leq d,\ x_i \wedge x_j > 0\}$. Therefore, the cone $\mathbb{C}_0$ consists of points of $\mathbb{C}$ that at most $d-2$ coordinates are 0.

Let F be a regularly varying distribution on $\mathbb{C}$, we say F possesses hidden regular variation if it is also regularly varying on $\mathbb{C}_0$. Then, there exits Radon measures v_α with tail index α, and v_{α_0} with index α_0 on $\mathbb{C}$ and $\mathbb{C}_0$ respectively, such that for scaling functions $b_0(t) \to \infty$ and $b(t)/b_0(t) \to \infty$ we obtain

$$t\mathbb{P}\left(\mathbf{b}\,(\mathbf{t})^{-\mathbf{1}}\,\mathbf{X} \in \cdot\right) \xrightarrow{v} v_\alpha(\cdot)$$

in $M_+(\mathbb{C})$ and

$$t\mathbb{P}\left(\mathbf{b_0}\,(\mathbf{t})^{-\mathbf{1}}\,\mathbf{X} \in \cdot\right) \xrightarrow{v} v_{\alpha_0}(\cdot) \tag{6.3.12}$$

in $M_+(\mathbb{C}_0)$.

The idea of this framework is to capture the delicate structure that may be present in the interior of the cone. For example, the multivariate normal distribution with correlation coefficient $|\rho| < 1$ of the Example 6.3.7 posses asymptotic independence, of the same form as a multivariate pareto distribution function with absolutely independent marginals. However, the structure of this asymptotic independence in both examples is very different and it cannot be describe by common measures of extreme dependence as the tail dependence function.

An equivalent form in polar coordinates for the hidden regular variation Condition 6.3.12 is

$$t\mathbb{P}\left(R\mathbf{b_0}\,(\mathbf{t})^{-\mathbf{1}}, \boldsymbol{\Theta} \in \cdot\right) \xrightarrow{v} c_0 v_{\alpha_0}(\cdot) \times S_0, \tag{6.3.13}$$

where $c_0 > 0$, S_0 is a Radon measure on $\mathbb{S}_0^{d-1} := \mathbb{S}^{d-1} \cap \mathbb{C}_0$ and convergence is in $M_+\left((0,\infty] \times \mathbb{S}_0^{d-1}\right)$.

Equation (6.3.13) shows that multivariate regular variation on $\mathbb{C}$ and hidden regular variation have the feature of being independent of a coordinates system. Some important consequences of this formulation are resumed in the next theorem.

Theorem 6.3.8. *Let be $b(\cdot) \in \mathcal{R}_{1/\alpha}$, $b_0(\cdot) \in \mathcal{R}_{1/\alpha_0}$, $b(t)/b_0(t) \to \infty$ and $0 < \alpha \leq \alpha_0$. Furthermore, assume*

$$t\mathbb{P}\left(\frac{X_j}{b(t)} > x\right) \to x^{-\alpha}\ ,\ j = 1, \ldots, d.$$

Then $\mathbf{X}$ possesses hidden regular variation iff

- *Regular variation on $\mathbb{C}$ implies $\mathbb{P}\left(\bigvee_{j=1}^d Z_j > x\right) \in \mathcal{R}_{-\alpha}$ while regular variation on $\mathbb{C}_0$ implies $\mathbb{P}\left(\bigwedge_{j=1}^d Z_j > x\right) \in \mathcal{R}_{-\alpha^0}$.*
- *Hidden regular variation implies asymptotic independence for $\mathbf{X} > 0$.*
- *Max-linear and min-linear combinations have regular varying tail probabilities with indeces α_j and α_0 respectively, this is*

$$t\mathbb{P}\left(\bigvee_{j=1}^{d} \frac{k_j X_j}{b(t)} > x\right) \to c(k)\, x^{-\alpha}$$

and

$$t\mathbb{P}\left(\bigwedge_{j=1}^{d} \frac{w_j X_j}{b_0(t)} > x\right) \to d(w)\, x^{-\alpha_0}$$

for $x > 0$, $t \to \infty$. *Where* $\mathbf{k} \in [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$, $c(\mathbf{k}) > 0$, $d(\mathbf{w}) > 0$ *and* $w \in (\mathbf{0}, \infty] \setminus \bigcup_{j=1}^{d} \left\{ t e_j^{-1}, t > 0 \right\}$ *with* $e_j^1 = \{\infty, \ldots, \infty, 1, \infty, \ldots, \infty\}$.

Having a finite hidden angular measure on $\mathbb{C}_0$ means that regular variation on $\mathbb{C}_0$ can be extended to regular variation of the same order on the full cone $\mathbb{C}$ and that marginals are regularly varying.

Maulik and Resnick (2004) showed how tail behaviour of the distribution F on $\mathbb{C}$ possessing hidden regular variation on $\mathbb{C}_0$ with finite hidden angular measure can be characterized in terms of tail equivalence. Conversely, they showed that if F is tail equivalent to a mixture distribution with $b(t)/b_0(t) \to \infty$, then F is multivariate regularly varying distribution on $\mathbb{C}$ and has hidden regular variation on $\mathbb{C}_0$ with finite spectral measure S_0.

The significance of having a finite angular measure on $\mathbb{C}_0$ implies that regular variation on $\mathbb{C}_0$ can be extended to regular variation of the same order on the full cone $\mathbb{C}$ and that marginals are regularly varying.

These ideas are resumed in the next Theorem.

THEOREM 6.3.9. *Let* F *be regularly varying on* $\mathbb{C}_0$ *with index* α_0, *scaling* $b_0(t)$, *limit measure* v_{α_0} *and spectral measure* S_0 *on* $\mathbb{S}_0^{d-1}$. *The following statemens are equivalent:*

(1) S_0 *is finite on* $\mathbb{S}_0^{d-1}$.

(2) *There exist a random vector* $F*$ *defined on* $\mathbb{C}$ *such that* $F* \overset{t.e\,\mathbb{C}^0}{\sim} F$, *where* $\overset{t.e\,\mathbb{C}^0}{\sim}$ *stays for tail equivalence on* $\mathbb{C}_0$, *and* $t\mathbb{P}(F*_j > b_0(t)x) \to cx^{-\alpha_0}$ *for* $j = 1, \ldots, d$ *as* $t \to \infty$, *for some* $c > 0$, *so that each component* $F*_j$ *has regularly varying tail probabilities with index* α_0.

(3) *There exists a random vector* $F*$ *defined on* $\mathbb{C}$ *such that* $F* \overset{t.e\,C^0}{\sim} F$, *such that for any* $\mathbf{k} \in [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$, *and any* $w \in (\mathbf{0}, \infty] \setminus \bigcup_{j=1}^{d} \left\{ t e_j^{-1}, t > 0 \right\}$ *with* $e_j^{-1} = \{\infty, \ldots, \infty, 1, \infty, \ldots, \infty\}$ *we have* $\bigvee_{j=1}^{d} k_j F*_j$ *and* $\bigwedge_{j=1}^{d} w_j F*_j$ *are tail equivalent* $\mathbb{C}_0$ *and have regular varying tail probabilities of index* α_0.

PROOF. See Maulik and Resnick (2004). $\square$

We need to define a standard form for hidden regular variation to be consistent with our framework.

DEFINITION 6.3.10. (Standard form of Hidden regular variation). Let $\mathbf{X}$ be a nonnegative random vector with regular varying distribution function $F \in \mathcal{R}_\alpha$. We say that F has hidden regular variation in standard form if in addition to Definition 6.3.5 there exists $b_0(t) \in \mathcal{R}_{1/\alpha_0}$ with $b_0(t) \to \infty$, $\alpha_0 \geq 1$, and

$$\lim_{t \to \infty} t/b_0(t) = \infty$$

such that

(1) $t\mathbb{P}\left(\mathbf{b}(\mathbf{t})^{\leftarrow}(\mathbf{X})/b_0(t) \in \cdot\right) \xrightarrow{v} v_{\alpha_0}(\cdot)$ on $\mathbb{C}_0$ and it is homogeneous of grade $-\alpha_0$, i.e., $v_{\alpha_0}(t\cdot) = t^{-\alpha_0} v_{\alpha_0}(\cdot)$.

(2) $t\mathbb{P}\left(\mathbf{b}(b_0(t))^{-1}\mathbf{X} \in \cdot\right) \xrightarrow{v} \mu(\cdot)$ on $\mathbb{C}_0$, where the Radon measures are related by $v_{\alpha_0}((\mathbf{x}, \infty]) = \mu((\mathbf{x}^\alpha, \infty])$ for all $x_j > 0$.

In addition Maulik and Resnick (2004) has shown that v_{α_0} can be either finite or infinite.[3] The situation where the hidden angular measure is infinite is very different and it has to be treated carefully. The idea is to take a compact subset of $\mathbb{C}_0$ and, hence, it will always have finite hidden measure. An useful subset for the infinite case is

$$\mathbb{S}_{inv}^{d-1} := \left\{ \mathbf{x} \in \mathbb{C}_0 : \bigwedge_{j=1}^{d} x_j \geq 1 \right\},$$

where we have to choose $b_0(t)$ such that $\mu_1^0 \left(\mathbb{S}_{inv}^{d-1} \right) = 1$.

It allows a standard form as finite as infinite hidden angular measure.

Example 6.3.11. Let X_1, X_2 be two iid Frèchet random variables with bivariate normal dependence with correlation $|\rho| < 1$ and that the standard marginal condition

$$t\mathbb{P}(X_i > b(t)) \sim 1 \tag{6.3.14}$$

is satisfied.

Ledford and Tawn (1996) have shown that if X_1, X_2 possesses hidden regular variation, then they admit the following representation

$$\mathbb{P}(X_1 \wedge X_2 > t) \sim \mathcal{L}(t)\,\mathbb{P}(X_1 > t)^{\alpha_0}$$

or equivalently by equation (6.3.14)

$$\mathbb{P}(X_1 \wedge X_2 > t) \sim \mathcal{L}'(t)\, t^{-\alpha_0},$$

where $\mathcal{L}, \mathcal{L}' \in \mathcal{R}_0$.

On the other hand, Reiss (1989, Chapter 7.) has shown that the joint survivor function at the same time satisfies

$$\mathbb{P}(X_1 \wedge X_2 > t) \sim (4\pi \log t)^{-\rho/(1+\rho)} \sqrt{\frac{(1+\rho)^3}{(1-\rho)}}\, t^{-2/(1+\rho)}, \tag{6.3.15}$$

as $t \to \infty$ for $\rho < 1$. Hence it is necessary that the hidden regular varying tail index is $\alpha_0 = \frac{2}{(1+\rho)}$.

Until now we are only introducing the main concepts in multivariate regular variation. Therefore, we still want to consider two topics: the relation among the correlation function, the tail of regular varying distribution and von Mises type functions and the estimation of fair away measure of probability in the case of asymptotic dependence and independence. The next section is dealing with these topics.

6.4. Implications of regular variation and hidden regular variation in risk management

The subject of this section is the implication of assuming conditional correlations with some classes of distributions as measure of extreme dependence to test contagion or interdependence. In particular, we use standard concepts in regular variation in the univariate

[3]See Maulik and Resnick (2004) for important results in this issue.

case to demonstrate that the correlation function is a measure depending on the marginal distribution, hence this is not suitable as measure of dependence for extreme events.

Second, we consider a framework to measure extreme spillover from the point of view of multivariate regular variation for the asymptotic dependent and independent cases. Furthermore, we give necessary and sufficient conditions to distinguish between the two cases.

6.4.1. Conditional variances in a Regular varying framework and its influence in factor models and models of contagion. The apparent increase in comovements of asset prices across markets during periods of financial turmoil has lead analysts and market observators to raise the hypothesis of "contagion" in the international transmission of currency and financial crises. Contagion is defined as a shock to one country's asset market that causes changes in asset prices in another country's financial market, whenever this increase is not significant, the phenomenon is seen as interdependence.

The major empirical papers[4] are using this approach found that there was a statistically significant increase in cross-market correlation coefficients during the relevant crises and therefore concluded that contagion occurred in the tested periods. The initial assumptions of this model are that the rates of return of the stock markets in two countries are linearly related. However, interpreting any increase in cross-country covariances and/or correlations as evidence of contagion, may be misleading.

This point can be illustrated by means of a simple example drawing on Forbes (2002). Suppose that the rates of return of the stock market in two countries are linearly related as

$$x_i = \lambda_i w + u_i$$

for $i = 1, 2$, where w_t is a random vector that represents common shocks, market fundamentals or non-diversifiable risk, which impact upon all asset returns with weights λ_i, while the random vectors u_i are idiosyncratic factors that are unique to a specific asset market. These returns could be on bond markets, equity markets, etc.

These models have the advantage that they provide convenient expressions for the conditional correlations between asset returns. Moreover, if the correlation between two time series is constant or is changing over time, one could consider comparing sampling correlations between the two series calculated from subsets of the data. Thus, if these conditional correlations are found to be statistically different from each other, one might be tempted to conclude that the population correlation is not constant. Following, we show analytically that this intuitively attractive approach to testing for correlation breakdowns can be very dubious.

Assume that we are interested in the correlation between x_i and w, and between x_1 and x_2, conditioned to the common factor $w \in \Omega$ where Ω is a convenient Borel set. Based on this assumption and in independence between $x_i$and w it is easy to derive

$$\text{Cov}\left(x_1, x_2 \mid w \in \Omega\right) = \lambda_1 \lambda_2 \text{Var}\left(w \mid w \in \Omega\right)$$

[4]Early works by Baig and Goldfajn (1999); Forbes (2002),Eichengreen et al. (1998); Calvo and Reinhart (1996) concentrate on this issue, while Corsetti et al. (2005); Caporale et al. (2003) are associated with a heteroskedasticity-adjusted correlation test that famously finds little evidence of contagion during a number of financial crises.

and

$$\operatorname{Var}(x_i \mid w \in \Omega) = \lambda_i^2 \operatorname{Var}(w \mid w \in \Omega) + \operatorname{Var}(u_i). \tag{6.4.1}$$

Therefore, the coditional correlations between $x_i$and w conditioned in $w \in \Omega$ can be written as

$$\rho(x_i, w \mid w \in \Omega) = \frac{\rho}{\sqrt{\rho^2 + (1-\rho^2)\frac{\operatorname{Var}(w)}{\operatorname{Var}(w \mid w\in\Omega)}}}, \tag{6.4.2}$$

where $\rho = \rho(x_i, w) = \frac{\lambda_i \operatorname{Var}(w)}{\sqrt{\lambda_i^2 \operatorname{Var}(w)^2 + \operatorname{Var}(u_i)}}$ is the unconditional correlation and

$$\rho(x_1, x_2 \mid w \in \Omega) = \frac{\lambda_1 \lambda_2}{\sqrt{\lambda_1^2 + \frac{\operatorname{Var}(u_1)}{\operatorname{Var}(w \mid w\in\Omega)}}\sqrt{\lambda_2^2 + \frac{\operatorname{Var}(u_2)}{\operatorname{Var}(w \mid w\in\Omega)}}} \tag{6.4.3}$$

is the correlation between x_1 and x_2 conditioned in the common factor $w \in \Omega$. Thus, both the variance (6.4.1) and the correlations (6.4.3) of x_i are functions of the conditional variance of w. Hence, if we assume a linear factor model for asset returns the diversification fail when $\operatorname{Var}(w \mid w \in \Omega) \to \infty$, exactly when we need it most.

Notice that $\rho(x_i, w \mid w \in \Omega)$ has the same sign that $\rho(x_i, w)$ and also that $\rho(x_i, w \mid w \in \Omega) \gtrless \rho(x_i, w)$ since $\operatorname{Var}(w \mid w \in \Omega) \gtrless \operatorname{Var}(w)$.

Based on this result, Boyer et al. (1999); Forbes and Rigobón (2001); Forbes (2002) between others, propose an adjust for this bias through a constant ratio $(1+\varepsilon)$ between the conditional variance of asset returns in tranquil(Ω_T) and crisis (Ω_C) periods

$$\operatorname{Var}(w \mid w \in \Omega_C) = (1+\varepsilon) \operatorname{Var}(w \mid w \in \Omega_T). \tag{6.4.4}$$

The justification of the idea in (6.4.4) is natural. We should expect the variance to be larger for those observations that fall into the tails of the distribution, simply because the variances of the tail observations are wider than the observations concentrates on the tranquil periods. Therefore, we would also expect the conditional correlation to be higher when w is in the tail of its distribution, irrespective of kind of tail distribution. However, this linear assumption for changes in correlations is not as straightforward as one might think.

6.4.1.1. *Conditional variance for Von Mises type and regular varying distributions.* Here we distinguish between Regular varying and Von Mises type distributions, instead of only light- and heavy tailed distributions, because of the fact that some heavy tailed distributions are not regular varying distributions as the Lognormal distribution. We haven already given a definition for regular varying distributions in Definition 6.2.9, therefore we adopted the below definition of Von Mises type distributions for distributions whose tails can be represented as a Von Mises function in Definition 6.4.1.

Definition 6.4.1. (Von Mises function, (Embrechts et al. (1997, Definition 3.3.18, pp. 138.)) Let F be a distribution function with right endpoint $x_F \leq \infty$. Suposse there exists some $z < x_F$ such that F has representation

$$\overline{F}(x) = c \exp\left\{-\int_z^x \frac{1}{a(t)} dt\right\}, \quad z < x \leq x_F,$$

where c is some positive constant, $a(\cdot)$ is a positive and absolutely continuos function with density a' and $\lim_{x \uparrow x_F} a'(x) = 0$. Then F is called a von Mises distribution function.

This class of functions belongs to the MDA of the Gumbel distribution. Furthermore, with some slight modifications of this definition yields a complete characterisation of the MDA of the Gumbel distribution (see Embrechts et al. (1997, pp. 142.)). Some further important results are resumed in the next Corollary. The proof of these results are for instance to be found in Embrechts et al. (1997, Chapter 3).

COROLLARY 6.4.2. *Assume that x is random vector . Then,*

- *if x has infinite right point $x_F = \infty$, then $\overline{F} \in \mathcal{R}_{-\infty}$. In particular, $E\left(X^+\right)^\alpha < \infty$ for all $\alpha > 0$, where $X^+ = \max(0, X)$ and $\lim_{x\to\infty} xf(x)/\overline{F}(x) = \infty$.*
- *if x has right point $x_F < \infty$, then $\overline{F}\left(x_F - x^{-1}\right) \in \mathcal{R}_{-\infty}$ and $\lim_{x\to x_F}(x_F - x) f(x)/\overline{F}(x) = \infty$.*
- *if x has right point $x_F \le \infty$, then the auxiliar function is $a(x) = \overline{F}(x)/f(x)$ if and only if $\lim_{x\uparrow x_F} \overline{F}(x) F''(x)/f(x)^2 = -1$.*

Some examples of Von Mises type distributions with their respective auxiliary functions $a(x)$ are:

Normal	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x > 0$, $x \in \mathbb{R}$,	$a(x) \sim x^{-1}$
Log normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{\{\ln x - \mu\}^2}{2\sigma^2}}$, $x > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$,	$a(x) \sim \frac{\sigma^2 x}{\ln x - \mu}$, $x \to \infty$.
Exponential like	$\overline{F}(x) \sim Be^{-\lambda x}$, $B, \lambda > 0$,	$a(x) \sim \lambda^{-1}$
Weibul like	$\overline{F}(x) \sim e^{-cx^\tau}$, $c, \tau > 0$, $x \ge 0$,	$a(x) = x^{1-\tau}/c\tau$, $x \ge 0$

The characterisation of tail distributions allow to provide the main result of this section.

PROPOSITION 6.4.3. *Let w be a random variable whose second moment exist and define the Borel set $\Omega = [t, \infty]$ for a high threshold $t \gg 0$. Then,*

(1) *If the distribution function of w satisfies Definition 6.4.1 with $a(w) = \overline{F}(w)/f(w)$, then $Var(w \mid w \in \Omega)$ is $a(w)^2$.*
(2) *If the distribution function of w satisfies Definition 6.4.1 and the tail distribution can be written as $\overline{F}(w) = a\exp(-bw)$ for $a, b > 0$, then $Var(w \mid w \in \Omega)$ is b^{-2}.*
(3) *if w has a distribution function $F(w) \in \mathcal{R}_\alpha$ for $\alpha > 2$. Then, $Var(w \mid w \in \Omega)$ is $\frac{t^2\alpha}{(\alpha-2)(\alpha-1)^2}$.*

Proposition 6.4.3 allows us to gain information about the behaviour of the conditional variance for very different types of distributions and to deduce some properties in the following corollary.

COROLLARY 6.4.4. *Let w be a random variable whose second moment there exist and $\Omega_1 = [t_1, \infty]$ and $\Omega_2 = [t_2, \infty]$ are two Borel set with $t_2 > t_1 > 1$. Then*

(1) *If the distribution function of w is one von Mises type, then $Var(w \mid w \in \Omega)/Var(w) \to \infty$ if $a(t) \to \infty$, or $Var(w \mid w \in \Omega)/Var(w) \to 0$ if $a(t) \to 0$, as $t \to \infty$.*
(2) *If the distribution function of w satisfies Definition 6.4.1 and the tail distribution can be written as $\overline{F}(w) = a\exp(-bw)$ for $a, b > 0$, then $Var(w \mid w \in \Omega_2) \sim Var(w \mid w \in \Omega_1)$, which means that w is independent of the conditional set Ω_1 and Ω_2.*
(3) *if w has a distribution function $F(w) \in \mathcal{R}_\alpha$ for $\alpha > 2$. Then, $Var(w \mid w \in \Omega_2) \sim Var(w \mid w \in \Omega_1)(t_2/t_1)^2$ as $t_1 \to \infty$.*

Notice in 2 that $\overline{F}(w) = a\exp(-bw)$ belongs to the class of Von Mises type distributions. However, this deserves special attention due to the behaviour of the truncated variance. This is because these light-tailed distributions have truncated variance independent of the truncation level, i.e. memoryless property. Hence, these distributions are not a suitable candidate to model assets returns.

Following we calculate some examples for these three cases but directly and not through an asymptotic estimation.

EXAMPLE 6.4.5. Let w_t be a random vector with exponential distribution where $\overline{F}(w) = \exp(-\lambda w)$, for $w_t \geq 0$ and $\lambda > 0$ is a von Mises function with auxiliar function $a(x) = 1/\lambda$, then the conditional variance of w in a Borel set $\Omega = [t, \infty]$ is given by

$$\begin{aligned}
\operatorname{Var}(w \mid w \in \Omega) &= \frac{\int_\Omega w^2 f(w)\,dw}{\overline{F}(w)} - \left(\frac{\int_\Omega wf(w)\,dw}{\overline{F}(w)}\right)^2 \\
&= \frac{\int_t^\infty w^2 f(w)\,dw}{\overline{F}(w)} - \left(\frac{\int_t^\infty wf(w)\,dw}{\overline{F}(w)}\right)^2 \\
&= \frac{t^2\exp(-\lambda t) + \lambda^{-1}2t\exp(-\lambda t) - \lambda^{-2}2\exp(-\lambda t)}{\exp(-\lambda t)} \\
&\quad - \left(\frac{t\exp(-\lambda t) + \lambda^{-1}\exp(-\lambda t)}{\exp(-\lambda t)}\right)^{-2} \\
&= \lambda^{-2}.
\end{aligned}$$

EXAMPLE 6.4.6. Let w_t be a random vector with normal distribution $\Phi(w)$ with von Mises function $a(x) \sim x^{-1}$. Notice that taking derivate to the normal density function $\varphi(w)$ we have $\varphi'(w) = -w\varphi(w)$ and $\varphi''(w) = -\varphi(w) - w\varphi'(w)$. Hence $\int_\Omega w\varphi(w)\,dw = -\Phi(w)$ and $\int_\Omega w^2\varphi(w)\,dw = -\int_\Omega w\varphi(w)\,dw = \varphi'(w) + \Phi(w)$.

Furthermore, $\Phi(w)$ can be expressed in terms of the complementary error function

$$\operatorname{erfc}(w) = \frac{2}{\sqrt{\pi}}\int_w^\infty \exp\left(-t^{-2}\right)dt$$

as $\Phi(w) = 1 - \frac{1}{2}\operatorname{erfc}\left(\frac{w}{\sqrt{2}}\right)$, $w \in \mathbb{R}$. Then the conditional variance of w in a Borel set $\Omega = [t, \infty]$ is given by

$$\begin{aligned}
\operatorname{Var}(w_t \mid w_t \in \Omega) &= \frac{\int_\Omega w^2 f(w)\,dw}{\overline{F}(w)} - \left(\frac{\int_\Omega wf(w)\,dw}{\overline{F}(w)}\right)^2 \\
&= \frac{\int_t^\infty w^2\varphi(w)\,dw}{\overline{\Phi}(t)} - \left(\frac{\int_t^\infty w\varphi(w)\,dw}{\overline{\Phi}(t)}\right)^2 \\
&= \frac{\varphi'(w)\mid_t^\infty + \overline{\Phi}(w)}{\overline{\Phi}(t)} - \left(\frac{-\varphi(w)\mid_t^\infty}{\overline{\Phi}(t)}\right)^2 \\
&= \frac{t\exp\left(-t^{-2}/2\right)/\sqrt{2\pi}}{\frac{1}{2}\operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right)} + 1 - \left(\frac{\exp\left(-t^{-2}/2\right)/\sqrt{2\pi}}{\frac{1}{2}\operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right)}\right)^2.
\end{aligned}$$

An useful asymptotic expansion of the complementary error function for large values of t

$$\operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right) = \frac{\sqrt{2}\exp\left(t^{-2}/2\right)}{\sqrt{\pi}t}\sum_{n=0}^{\infty}(-1)^n \frac{(2n)!}{2!\left(\sqrt{2}t\right)^{2n}},$$

which yields to

$$\operatorname{Var}(w \mid w \in \Omega) \quad \sim \quad t^{-2} + \mathcal{O}\left(t^{-4}\right).$$

EXAMPLE 6.4.7. Let w be a random vector with pareto distribution $\overline{F} \in \mathcal{R}_{-\alpha}$ with parameter $\alpha > 2$ and density $f \in \mathcal{R}_{-(1+\alpha)}$. Then the conditional variance of w_t in a Borel set Ω is derived easily applying Karamatas representation (see Corollary (6.2.5)).

$$\begin{aligned}\operatorname{Var}(w_t \mid w_t \in \Omega) &= \frac{\int_\Omega w^2 f(w)\,dw}{\overline{F}(w)} - \left(\frac{\int_\Omega w f(w)\,dw}{\overline{F}(w)}\right)^2 \\ &= \frac{t^2\alpha}{(2-\alpha)(1-\alpha)^2}.\end{aligned}$$

EXAMPLE 6.4.8. Let w be a random vector with Student's t-distribution F, $v > 2$ degree of freedom and density

$$f(w) = t_v(w) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}}\left(1 + \frac{w^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$$

Let us define $\triangle_v = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}}$ and $\overline{F}(w) = \overline{T}_v(w) = v^{\frac{v-1}{2}}\frac{\triangle_v}{w^v} + \mathcal{O}\left(w^{-(v+2)}\right)$. Then the conditional variance of w in a Borel set $\Omega = [t, \infty]$ is given by

$$\begin{aligned}\operatorname{Var}(w \mid w \in \Omega) &= \frac{\int_\Omega w^2 f(w)\,dw}{\overline{F}(w)} - \left(\frac{\int_\Omega w f(w)\,dw}{\overline{F}(w)}\right)^2 \\ &= v\left[\frac{(v-1)\,\overline{T}_{v-2}\left(\sqrt{\frac{v-2}{v}}w\right)}{(v-2)\,\overline{T}_v(w)} - 1\right] - \left(\frac{\sqrt{\frac{v-2}{v}}t_v\left(\sqrt{\frac{v-2}{v}}w\right)}{\overline{T}_v(w)}\right)^2 \\ &\sim \frac{v}{(v-2)(v-1)^2}w^2 + \mathcal{O}(1).\end{aligned}$$

Hereafter, we present the asymptotic expression of $\rho(x_i, w \mid w \in \Omega)$ for the random variable w in the linear factor model (6.4.2).

COROLLARY 6.4.9. *Let us assume the linear factor model $x_i = \lambda_i w + u_i$ with unconditional correlation $\rho(x_i, w) = \frac{\lambda_i Var(w)}{\sqrt{\lambda_i^2 Var(w)^2 + Var(u_i)}}$, then*

(1) *If the distribution function of $w \sim N(0,1)$, then $\rho(x_i, w \mid w \in \Omega) \sim \frac{\rho}{t\sqrt{(1-\rho^2)}}$ for $t \to \infty$.*
(2) *If the distribution function of $w \sim LN(0,1)$, then $\rho(x_i, w \mid w \in \Omega) \sim \frac{\rho \ln t}{t\sqrt{(1-\rho^2)}}$ for $t \to \infty$.*
(3) *If the distribution function of $w \sim \exp(\lambda)$, then $\rho(x_i, w \mid w \in \Omega) = \rho(x_i, w)$ for all $t > 0$.*
(4) *If the distribution function of $w \sim Weibull(c, \tau, \lambda)$, then $\rho(x_i, w \mid w \in \Omega) \sim \frac{\rho c\tau}{t^{(1-\tau)}\sqrt{(1-\rho^2)}}$ for $t \to \infty$.*

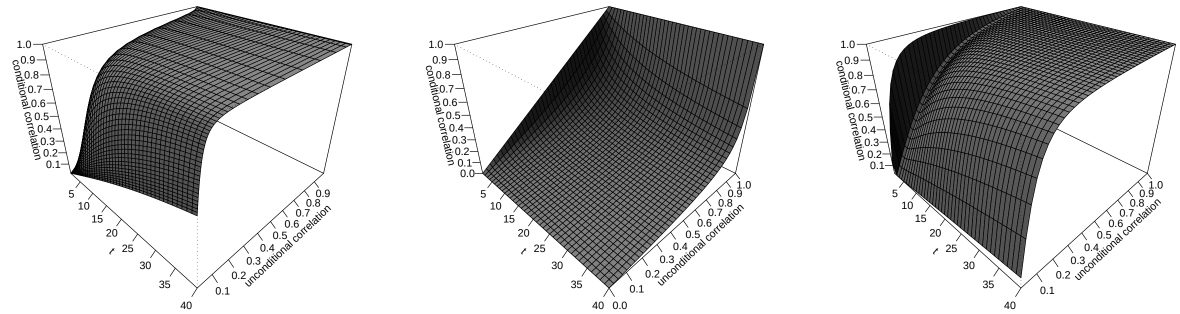

FIGURE 6.4.1. Theoretical conditional correlation $\rho(x_i, w \mid w \in \Omega)$ and unconditional correlation $\rho(x_i, w)$ in function of a subset $\Omega = [t, \infty]$, i.e., for high threshold t, When $\overline{F}(w) \in \mathcal{R}_{-\alpha}$ (left), $w \sim N(0,1)$(middle) and $w \sim LN(0,1)$ (right).

(5) *If the distribution function of w is regular vaying with index $\alpha > 2$, then $\rho(x_i, w \mid w \in \Omega) \sim \frac{\rho t}{(\alpha-1)\sqrt{(1-\rho^2)}}$ for $t \to \infty$.*

Figure 6.4.1 illustrates the theoretical relationship between the conditional correlation, the unconditional correlation and conditioning events for three types of distributions, regular varying distributions for $\overline{F}(w) \in \mathcal{R}_{-\alpha}$, a normal distribution $w \sim N(0,1)$ and a Log-Normal distribution $w \sim LN(0,1)$. In this example we assume that u_i is Normal distributed with zero mean and variance equal to the unity. As it is shown in the case of regular varying functions, the conditional correlation converges very fast to the limit one for high values of the threshold t and it does not depend strongly on the chosen unconditional correlation.

On the contrary, the results for the tail distributions of von Mises type are varied. For the normal distribution the conditional correlation tends very slow to zero for high threshold of the conditional set, while a stronger relationship with unconditional correlation is observed. In the case of the Log-Normal distribution the conditional correlation very slowly tends to one for high threshold of the conditional set, while when $t \downarrow 1$ we find a breakdown produced by the properties of the marginal distribution.

From the preceding discussion we note that knowledge of the tail distribution behaviour lets us determine whether the conditional variance is less than or greater than the unconditional variance and therefore the correlation function. Further, the empirical illustrations confirm also that it would be improper to conclude that the correlation between two series varies across observations based on sample-splitting exercises alone as was done for example in (Caporale et al. (2003); Corsetti et al. (2005); Dungey et al. (2005)).

On the other hand and for completeness the next examples illustrates that even having correlation $\rho(x_i, w) = 0$ in a linear factor model the true dependence between x_i and w can be very different.

EXAMPLE 6.4.10. Define the following random variables $w \sim LN(0,1)$ and $x_i \sim LN(0, \sigma)$, and joint distribution $\mathcal{W}(x_i, w)$. We know that all bivariate distribution function have as

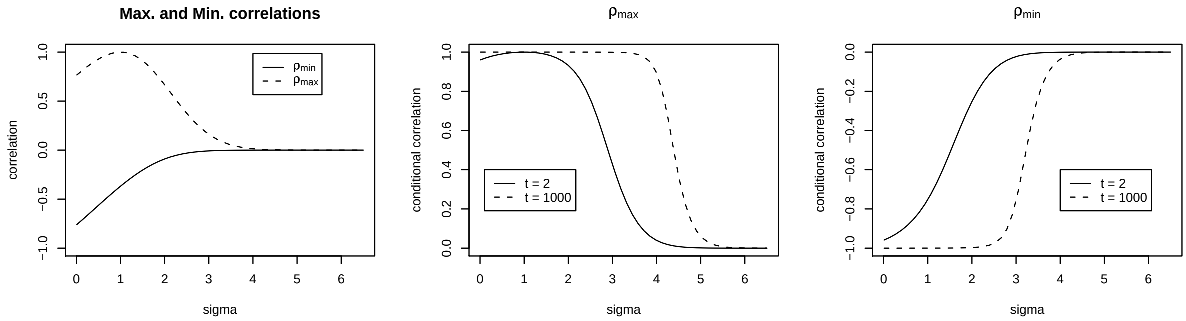

FIGURE 6.4.2. Left :Maximum and minimum attainable correlation for different values of σ in example (6.4.10).
Middle: Conditional correlation for the maxima attainable correlation ($\rho_{\max}$) for two conditional thresholds ($t = 2$ and $t = 1000$).
Rigth: Conditional correlation for the minimum attainable correlation ($\rho_{\min}$) for two conditional threshold ($t = 2$ and $t = 1000$).

Frèchet bounds

$$\max\left\{\mathcal{W}_1(x_i) + \mathcal{W}_2(w) - 1, 0\right\} \leq \mathcal{W}(x_i, w) \leq \min\left\{\mathcal{W}_1(x_i), \mathcal{W}_2(w)\right\}$$

Clearly, when the marginals distribution $\mathcal{W}_1(x_i)$ and $\mathcal{W}_2(w)$ are fixed, the correlation has to be between the minima and maxima correlation of these Frèchet bounds (see McNeil et al. (2005, Chapter 5)). In fact, for our marginal distributions we have that

$$\rho_{\min}(x_i, w) = \frac{e^{-\sigma} - 1}{\sqrt{(e-1)\left(e^{\sigma^2} - 1\right)}}, \qquad \rho_{\max}(x_i, w) = \frac{e^{\sigma} - 1}{\sqrt{(e-1)\left(e^{\sigma^2} - 1\right)}}.$$

This shows as σ increases, one notes how the boundaries of the interval $(\rho_{\min}, \rho_{\max})$ tend rapidly to zero and therefore, for any attainable conditional set $\Omega = [t, \infty]$, the conditional correlation $\rho(x_i, w \mid w \in \Omega) \to 0$. Moreover, since $\mathcal{W}(x_i, w)$ experiments the strongest form of positive dependence when x_i and w are comonotonic, i .e., $\min\left\{\mathcal{W}_1(x_i), \mathcal{W}_2(w)\right\}$, this results illustrates as small correlations not necessarily imply weak dependence.

The message of the last example is that any interpretation of correlation is meaningless without knowing the true dependence or copula of a joint model.

The central argument of this chapter is that computing correlations condition on realization of one variable and observing that those correlations are different for miscellaneous conditioning events, gives you no basis to conclude that the true dependence of the data-generating process is changing over time. Furthermore, this section showed as the correlation between two variables conditioned on a high subset of exceedances may deviate significantly from the unconditional correlation.

One could hope for the existence of logical links between some of these measures, such as a vanishing tail dependence parameter implies vanishing asymptotic conditional correlation coefficients. Indeed, this turns out to be wrong and one can construct simple examples for which all possible combinations occur as in example 6.4.10.

The conditional correlation is sensitive to the marginals, i.e., it is a measure of extreme dependence weighted by the specific shapes of the marginals, while the dependence measures introduced in Section 2.4 are a pure copula property independent of the marginals. When conditioning on w, one changes the copula between x_i and w, so that the extreme dependence properties by the conditional correlations are not exactly those of the original copula. Indeed, we can find situations in which we are in the presence of asymptotic independence, while the conditional correlation could conclude the opposite.

These results demonstrate the difficulty and the controversy that can be the results of the existence of contagion or interdependence in financial crisis, when we define these in terms of conditional correlations. Furthermore, even if the bias has been accounted for, as soon as the property of regular variation is present in the data with tail index $\alpha > 2$ for a sample size n, the nature of the convergence of the empirical Pearson's correlation coefficient of a bivariate time series is of the order of $n^{1-2/\alpha}$ for the error between the empirical and theoretical correlation. Moreover, the tail index has a stable law with index $\alpha/2$ (see Meerschaert and Scheffler (2001)).

6.4.2. Estimations of exponent measures. In the above sections we have seen that multivariate regular variation and hidden regular variation models are characterized principally by tail indices and exponent measures or alternatively a spectral measure. We attempt to do both here by estimating spectral measures, exceedence probabilities and some other features of the tails.

In practical application of these methods is normally suggested to use nonparametric models for estimating the dependence structure because of infinite number of possible parameterization.

In Subsection 6.4.2.1 we consider estimation of the exponent measure and the spectral measure in the standard case of multivariate regular variation. Further, in Section 6.4.2.2 we shall discuss a simple method that allows for a more precise estimation of the hidden spectral regular measure and its corresponding hidden exponent measure. Finally, we discuss the relation between the second order regular variation and hidden regular variation.

6.4.2.1. *Estimation in the standard case the cone* $\mathbb{C} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$. Let $\mathbf{X}$ be a d-variate random vector whose common distribution F is multivariate regular varying. Recall that from Definition 6.3.2, multivariate regular variation is equivalent to the existence of $b(t) \to \infty$ such that the limit

$$t\mathbb{P}\left(\mathbf{X}b(t)^{-1} \in \cdot\right) \xrightarrow{v} v_\alpha(\cdot) \tag{6.4.5}$$

exists.

The estimation of this type the probabilities are ever-present in extreme value theory and the solution is to move estimation from the boundary of the sample to more safer areas with more observations. The idea is to find a sequence $k \to \infty$ with $k/n \to \infty$ such that limit (6.4.5) still holds, when we substitute t with n/k.

$$(n/k)\mathbb{P}\left(\mathbf{X}b(n/k)^{-1} \in \cdot\right) \xrightarrow{v} v_\alpha(\cdot)\,. \tag{6.4.6}$$

Now the question is how to find an empirical measure to estimate such class of limit probabilities. An answer to this question is provided by Resnick (2006, page 179) who shows that multivariate regular variation of the probability distributions is equivalent to induced empirical measures weakly converging to Poisson random measure limits (Definition 6.3.2, point 6). Thus,

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\mathbf{X}b(n/k)^{-1}} \Rightarrow v_{\alpha}$$

is a limit measure for known $b(k/n)$. However, this measure assumes that the tails are equivalent. In practice, it is unusual to conclude that the tail index α_j is the same for each component. Thus, a standardized form is necessary.

The idea is that if $\mathbf{X}$ is a d-variate random vector whose tails are asymptotically Pareto, that is, for each marginal $j = 1, \ldots, d$, $\mathbb{P}(X_j > x_j) \sim x_j^{-\alpha_j}$, then, it implies that $\mathbb{P}\left(X_j^{\alpha} > x_j\right) \sim x_j^{-1}$, for $x_j \to \infty$. Hence,

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left\{(X_j/b_j(n/k))^{\alpha_j};\, j=1,\ldots,d\right\}} \Rightarrow \mu\left([\mathbf{0}, \mathbf{x}^{\alpha}]^c\right)$$

is a standard limit measure, whose marginals are all Pareto with homogeneous degree -1, that is,

$$\mu(t\cdot) = t^{-1}\mu(\cdot)\,. \tag{6.4.7}$$

The above results can also be expressed in terms of polar coordinates. In the next proposition we summarize.

Proposition 6.4.11. *Let $\mathbf{X}$ be a d-variate random vector whose common distribution F is multivariate regular varying, with marginal indeces α_j and sequence $b_j(n/k)$ for $j = 1, \ldots, d$ and $k \to \infty$ with $k/n \to \infty$. Then,*

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left\{(X_{ij}/b_j(n/k))^{\alpha_j};\, j=1,\ldots,d\right\}} \tag{6.4.8}$$

converges to the standard limit measure μ in $M_+(\mathbb{C})$ or equivalently in polar coordinates

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(R_i, \Theta_{i,k}\right)} \Rightarrow c\mu \times \mathcal{S} \quad \text{in } M_+\left((0, \infty] \times \mathbb{S}_+^{d-1}\right), \tag{6.4.9}$$

where $c > 0$ and $\mathcal{S}$ is a probability measure on Borel subsets on $\mathbb{S}_+^{d-1}$.

Since we want to estimate the probability of a set beyond the sample in the tail, we will use the homogeneous property (6.4.7) to estimate this probability. Instead of using equations (6.4.8) or (6.4.9) diectly, we use a scaling parameter so that there are some sample points in the set to estimate. The procedure is as follows.

Assume that we are interested in a probability of the type

$$\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) \tag{6.4.10}$$

with $x_j \to \infty$ for all j and the Proposition 6.4.11 holds.

de Haan and de Ronde (1998) proposes the following approximation for the p-th quantile of the threshold $x_j = x_j^{(p)} \sim b_j(n/k)\left(k/n\left(1 - p_j\right)\right)^{1/\alpha_j}$. Making use of this approximation, the empirical estimator (6.4.8) and the scaling property in equation (6.4.10) we obtain

$$\begin{aligned}\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) &\sim \frac{k}{n}\frac{n}{k}\mathbb{P}\left(\bigcap_{j=1}^{d}\left\{X_j > b_j\,(n/k)\,(k/n\,(1-p_j))^{1/\alpha_j}\right\}\right)\\ &= \frac{k}{n}\mu_1\left(((k/n\,(1-p_j)\,;\,j=1,\ldots,d)\,,\infty]\right)\\ &= \frac{k}{n}\mu_1\left((k/n\,(\mathbf{1}-\mathbf{p})\,,\infty]\right).\end{aligned}$$

Define $\|\cdot\|$ as any arbitrary norm in $\mathbb{R}^d$, then applying the scaling $\|k/n\,(\mathbf{1}-\mathbf{p})\|^{-1}$

$$\begin{aligned}\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) &\sim \frac{k}{n\,\|k/n\,(\mathbf{1}-\mathbf{p})\|}v_\alpha\left(\left(\frac{k/n\,(\mathbf{1}-\mathbf{p})^\alpha}{\|k/n\,(\mathbf{1}-\mathbf{p})\|^\alpha},\infty\right]\right)\\ &= \frac{k}{n\,\|k/n\,(\mathbf{1}-\mathbf{p})\|}\mu\left(\left(\frac{k/n\,(\mathbf{1}-\mathbf{p})}{k/n\,(\mathbf{1}-\mathbf{p})},\infty\right]\right)\\ &= \left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|^{-1}\mu\left(\left(\left(\frac{(1-p_j)^{-1}}{\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|}\; j=1,\ldots,d\right),\infty\right]\right)\end{aligned}$$

replacing μ by the empirical measure as in Definition 6.3.2 we obtain

$$\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) \sim \frac{1}{\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|}\sum_{i=1}^{n}\epsilon_{\left\{(X_{ij}/b_j(n/k))^{\alpha_j}\in\left(\left(\frac{(1-p_j)^{-1}}{\|(\mathbf{1}-\mathbf{p})^{-1}\|};\,j=1,\ldots,d\right),\infty\right]\right\}}. \tag{6.4.11}$$

Note that this spillover probability is estimated on an infinite square whose lower left corner always falls on the unit circle[5], a statement that remains true irrespectively of how large the quantiles $p_1,\ldots,p_d$ are chosen. We provide an example to elucidate some results.

EXAMPLE 6.4.12. Consider $\mathbf{X}$ is a bivariate random vector whose common distribution F is multivariate regular varying, with marginal indices α_1, α_2 and sequence $b_j\,(n/k)$ for $j=1,2$. Furthermore, there exits $k\to\infty$ such that $k/n\to\infty$.

When we are interested and the conditional probability of the type $\mathbb{P}\,(X_1 > x_1 \mid X_2 > x_2)$ for $x_1, x_2\to\infty$, by the empirical measure of equation (6.4.11) we obtain:

$$\lim_{x_1,x_2\to\infty}\mathbb{P}\,(X_1 > x_1 \mid X_2 > x_2) = \lim_{x_1,x_2\to\infty}\frac{\mathbb{P}\,(X_1 > x_1,\, X_2 > x_2)}{\mathbb{P}\,(X_2 > x_2)}$$

$$\sim \frac{(1-p_2)^{-1}}{\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|}\sum_{i=1}^{n}\epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1},\left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right)\in\left(\left(\frac{(1-p_1)^{-1}}{\|(\mathbf{1}-\mathbf{p})^{-1}\|},\infty\right]\times\left(\frac{(1-p_1)^{-1}}{\|(\mathbf{1}-\mathbf{p})^{-1}\|},\infty\right]\right)\right\}}$$

define $(1-p_1)=\vartheta\,(1-p_2)$ and using the three norms $\|\cdot\|_{\text{sum}}$, $\|\cdot\|_{\max}$, $\|\cdot\|_2$ we get for the norm $\|\cdot\|_{\text{sum}}$

$$\left(\frac{\vartheta}{1+\vartheta}\right)\sum_{i=1}^{n}\epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1},\left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right)\in\left(\left(\frac{1}{1+\vartheta},\infty\right]\times\left(\frac{\vartheta}{1+\vartheta},\infty\right]\right)\right\}},$$

for the norm $\|\cdot\|_{\max}$,

[5] In the case of the Euclidean norm.

$$1 \wedge \frac{(1-p_1)}{(1-p_2)} \sum_{i=1}^{n} \epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1}, \left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right) \in \left(\left(\frac{(1-p_1)\wedge(1-p_2)}{(1-p_1)}, \infty\right] \times \left(\frac{(1-p_1)\wedge(1-p_2)}{(1-p_2)}, \infty\right]\right)\right\}}$$

for the euclidean norm $\|\cdot\|_2$

$$\left(\frac{\vartheta}{\sqrt{1+\vartheta}}\right) \sum_{i=1}^{n} \epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1}, \left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right) \in \left(\left(\frac{1}{\sqrt{1+\vartheta}}, \infty\right] \times \left(\frac{\vartheta}{\sqrt{1+\vartheta}}, \infty\right]\right)\right\}}.$$

In the case of the angular measure $\mathcal{S}$ we seek to transform first to the standard case with a standard limit measure μ and then to applythe polar coordinate transform $T(x) = \left(\|x\|, \frac{x}{\|x\|}\right) = (x, \theta)$ to get

$$T\left((X_{ij}/b_j(n/k))^{\alpha}, i = 1, \ldots, n\right) = \left((R_{i,k}, \Theta_{i,k})\,;\, i = 1, \ldots, n\right)$$

and from (6.4.9) we get the estimator

$$S_n(\cdot) = \frac{\sum_{i=1}^{n} \epsilon_{\{R_{ik}, \Theta_{ik}\}}\left((1,\infty] \times \cdot\right)}{\sum_{i=1}^{n} \epsilon_{\{R_{ik}, \Theta_{ik}\}}\left((1,\infty]\right)} \Rightarrow S(\cdot) \tag{6.4.12}$$

The interpretation of (6.4.17) is to chose Θ_{ik} whose radius is greater than 1. The angular measure is in this case defined on the interval $[0, \pi/2]$. For instance, if the most of the mass density concentrates around $\pi/4$ then the dependence between extreme movements of the marginals is stronger, while the most of the mass around 0 and $\pi/2$ shows a tendency toward independence , or at least asymptotic independence.

Until now in this section, we have supposedly known the threshold or equivalently the number of order statistics k, the tail indices of the marginals α_j and the parameter $b(n/k)$. The next proposition answers these questions.

PROPOSITION 6.4.13. *Consider* $\mathbf{X}$ *a d-variate random vector whose common distribution F is multivariate regular varying, with marginal indices α_j. Then,*

(1) $\hat{b}(n/k) = X_{k,n}$ *is a consistent estimator of* $b(n/k)$.
(2) *The estimator* $\hat{k} = \arg\left\{\frac{t\hat{\mu}_1(t\mathbb{S}^{d-1})}{\hat{\mu}_1(\mathbb{S}^{d-1})} = 1\right\}$, *with t varying in the neighborhood of 1, is a good estimator for k-th largest order statistic of the sample.*
(3) *The Hill estimator* $H^{(j)}_{\hat{k},n} := \frac{1}{\hat{k}} \sum_{i=1}^{\hat{k}} \log\left(X_{n,n-i+1}/X_{n,n-\hat{k}+1}\right) \to \alpha_j^{-1}$, *where $X_{n,k}$ is the k-th largest order statistic of the sample X_n, is a consistent estimator for the tail indeces α_j.*

6.4.2.2. *The hidden regular measure.* A first approximation to detect if existing hidden regular variation is suggested by the Theorem 6.3.8. First, each marginal has to be standardized to Pareto marginals with tail index $\alpha_j = 1$ by means of a power transformation. As it was explained in Proposition 6.4.11.

By Theorem 6.3.8 there exists evidence for Hidden regular variation if the minimum of the d-variate standardized marginals has tail index $\alpha_0 > 1$. Thus, a first approximation should be to plot the Hill estimator for the minimum of the standardized data. Assuming that (6.3.8) and (6.3.9) hold, then necessarily the probability $\mathbb{P}\left(\bigcap_{j=1}^{d}\left\{X_j b(t)^{-1} > x_j\right\}\right)$ has to be zero.

In fact, if I denotes a subset of the set D of marginals we obtain that

$$\begin{aligned} t\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j/b(t) > x_j\}\right) &= \sum_{I\subset D} |-1|^{|I|}\, t\mathbb{P}\left(\bigcup_{k\in I}^{d}\{X_k/b(t) > x_k\}\right) \qquad (6.4.13)\\ &\le \sum_{\substack{I\subset D\\ |I|>2}} t\mathbb{P}\left(\bigcap_{k\in I}^{d}\{X_k/b(t) > x_k\}\right)\\ &= \sum_{\substack{I\subset D\\ |I|>2}} t\mathbb{P}\left(\bigcap_{k\in I}^{d}\{X_k/b_0(t) > b(t)x_k/b_0(t)\}\right)\\ &= \sum_{\substack{I\subset D\\ |I|>2}} t\mathbb{P}\left(\bigwedge_{k\in I}^{d}\{\left(x_k^{-1}\right)X_k\}/b_0(t) > b(t)/b_0(t)\right)\\ &\to 0. \end{aligned}$$

Making use of Definition 6.3.10 we have that

$$t\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j/b_0(t) > x_j\}\right) \to \mathbf{x}^{-\alpha_0} v_{\alpha_0}\left((\mathbf{1},\infty]\right)$$

and by (6.4.13) note that

$$t\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j/b_0(t) > x_j\}\right) = t\mathbb{P}\left(\bigwedge_{j=1}^{d}\left\{\left(x_j^{-1}\right)X_j\right\}/b_0(t) > 1\right) \qquad (6.4.14)$$

for $0 < x_j < \infty$ and $j = 1, \ldots, d$.

Since that $b_0(t)$ is unknown for statistical purposes equation (6.4.14) and the fact $v_0\left(\mathbb{S}^{+}_{inv}\right) = 1$ give the next result, which allows to replaces $b_0(t)$ by a statistic.

PROPOSITION 6.4.14. *Consider* $\mathbf{X}$ *is a d-variate random vector whose common distribution* F *possesses both regular variation and hidden regular variation with marginal indeces* α_j *and* α_0. *Then, if* $W_i = \bigwedge_{j=1}^{d} X_{ij}$ *is random vector composed by the minima of the random vector* $\mathbf{X}$ *we obtain,*

(1) $\hat{b}_0(n/k) := W_{n,k}$ *is a consistent estimator of* $b_0(n/k)$, *where* $W_{n,k}$ *is the k-th largest order statistic of the sample* W_n.

(2) $\hat{v}_{\alpha_0} := \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(X_{ij}/\hat{b}_0(n/k),\, 1\le j\le d\right)} \Rightarrow v_{\alpha_0}$ *in* $M_+(\mathbb{C}_0)$.

Heffernan and Resnick (2005) have proven the last proposition using rank transformation. However, we are interested in estimate extreme probabilities of tail regions, where the number of events that have happened are even zero.

A more interesting approximation is the use of the idea proposed in the case of asymptotic dependence, where we move the region of extreme events so that the number of observations is larger. First, observe in Definition 6.3.10 that hidden regular measure v_{α_0} is related to the classical measure of regular variation μ by $v_{\alpha_0}((\mathbf{x},\infty]) = \mu((\mathbf{x}^{\alpha},\infty])$. Second, if we use the scaling property for the hidden measure to blow the sample we obtain

$$\mathbf{x}^{\alpha_0} v_{\alpha_0}((\mathbf{1},\infty]) = \mathbf{x}^{\alpha_0}\mu((\mathbf{1}^{\alpha},\infty]). \tag{6.4.15}$$

Now, as in the above section, suppose that we are interested in a probability of the type

$$\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right)$$

with $x_j \to \infty$ for all j and in this case the Definition 6.3.10 holds, that there exist hidden regular variation.

Proceeding as in the classical case of regular variation using the empirical estimator (6.4.8) and the scaling property in equation 6.4.15) we obtain

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) &\sim \frac{k}{n}\frac{n}{k}\mathbb{P}\left(\bigcap_{j=1}^{d}\left\{X_j > b_j(n/k)\left(k/n(1-p_j)\right)^{1/\alpha_j}\right\}\right)\\
&= \frac{k}{n}\mu\left(\left(\left(k/n(1-p_j)\,;\, j=1,\ldots,d\right),\infty\right]\right)\\
&= \frac{k}{n}\mu\left(\left(k/n(\mathbf{1}-\mathbf{p}),\infty\right]\right).
\end{aligned}$$

Define $\|\cdot\|$ as any norm in $\mathbb{R}^d$, then applying the scaling $\|k/n(\mathbf{1}-\mathbf{p})\|^{-1}$

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) &\sim \frac{k}{n\|k/n(\mathbf{1}-\mathbf{p})\|^{\alpha_0}}v_{\alpha_0}\left(\left(\mathbf{0},\frac{k/n(\mathbf{1}-\mathbf{p})}{\|k/n(\mathbf{1}-\mathbf{p})\|}\right]^{c}\right)\\
&= \frac{k}{n\|k/n(\mathbf{1}-\mathbf{p})\|^{\alpha_0}}\mu\left(\left(\mathbf{0},\left(\frac{k/n(1-p_j)}{\|k/n(\mathbf{1}-\mathbf{p})\|};\, j=1,\ldots,d\right)\right]^{c}\right)\\
&= \left(\frac{k}{n}\right)^{1-\alpha_0}\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|^{-\alpha_0}\mu\left(\left(\mathbf{0},\left(\frac{(1-p_j)^{-1}}{\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|}\; j=1,\ldots,d\right)\right]^{c}\right)
\end{aligned}$$

replacing μ by the empirical measure we obtain

$$\mathbb{P}\left(\bigcap_{j=1}^{d}\{X_j > x_j\}\right) \sim \left(\frac{k}{n}\right)^{1-\alpha_0}\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|^{-\alpha_0}\sum_{i=1}^{n}\epsilon_{\left\{(X_{ij}/b_j(n/k))^{\alpha_j}\in\left(\mathbf{0},\left(\frac{(1-p_j)^{-1}}{\|(\mathbf{1}-\mathbf{p})^{-1}\|};\, j=1,\ldots,d\right)\right]^{c}\right\}}. \tag{6.4.16}$$

We give an example to elucidate some results.

Example 6.4.15. Consider $\mathbf{X}$ a bivariate random vector whose common distribution F is multivariate regular varying, with marginal indeces α_1 and α_2 and standard regular varying index $\alpha = 1$ and hidden regular variation index α_0, for $k \to \infty$ with $k/n \to \infty$. We are interested and the conditional probability of the type $\mathbb{P}(X_1 > x_1 \mid X_2 > x_2)$ for $x_1, x_2 \to \infty$.

Applying the empirical measure of equation (6.4.16) we obtain

$$\lim_{x_1,x_2\to\infty} \mathbb{P}\left(X_1 > x_1 \mid X_2 > x_2\right) \quad = \lim\nolimits_{x_1,x_2\to\infty} \quad \frac{\mathbb{P}\left(X_1 > x_1,\ X_2 > x_2\right)}{\mathbb{P}\left(X_2 > x_2\right)}$$

$$\sim \frac{(1-p_2)}{\left\|(\mathbf{1}-\mathbf{p})^{-1}\right\|^{\alpha_0}} \sum_{i=1}^{n} \epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1}, \left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right) \in \left(\left(\frac{(1-p_1)^{-1}}{\|(\mathbf{1}-\mathbf{p})^{-1}\|},\infty\right] \times \left(\frac{(1-p_1)^{-1}}{\|(\mathbf{1}-\mathbf{p})^{-1}\|},\infty\right]\right)\right\}}$$

define $(1-p_1) = \vartheta\,(1-p_2)$ and using the three norms $\|\cdot\|_{\text{sum}}$, $\|\cdot\|_{\max}$, $\|\cdot\|_2$, we get for the norm $\|\cdot\|_{\text{sum}}$,

$$\left(\frac{k}{n\,(1-p_2)}\right)^{1-\alpha_0} \left(\frac{\vartheta}{1+\vartheta}\right)^{\alpha_0} \sum_{i=1}^{n} \epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1}, \left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right) \in \left(\left(\frac{1}{1+\vartheta},\infty\right] \times \left(\frac{\vartheta}{1+\vartheta},\infty\right]\right)\right\}}$$

for the norm $\|\cdot\|_{\max}$

$$\left(\frac{k}{n}\right)^{1-\alpha_0} \left(\frac{(1-p_2)^{\alpha_0} \wedge (1-p_1)^{\alpha_0}}{(1-p_2)}\right) \sum_{i=1}^{n} \epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1}, \left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right) \in \left(\left(\frac{(1-p_1)\wedge(1-p_2)}{(1-p_1)},\infty\right] \times \left(\frac{(1-p_1)\wedge(1-p_2)}{(1-p_2)},\infty\right]\right)\right\}}$$

and at the end for the euclidean norm $\|\cdot\|_2$

$$\left(\frac{k}{n\,(1-p_2)}\right)^{1-\alpha_0} \left(\frac{\vartheta}{\sqrt{1+\vartheta}}\right)^{\alpha_0} \sum_{i=1}^{n} \epsilon_{\left\{\left(\left(\frac{X_{i1}}{b_1(n/k)}\right)^{\alpha_1}, \left(\frac{X_{i2}}{b_2(n/k)}\right)^{\alpha_2}\right) \in \left(\left(\frac{1}{\sqrt{1+\vartheta}},\infty\right] \times \left(\frac{\vartheta}{\sqrt{1+\vartheta}},\infty\right]\right)\right\}}.$$

In the case of the angular measure $\mathcal{S}$ we seek to transform to the standard case with a standard limit measure μ. Then, we apply the polar coordinate transformation $T(x) = \left(\|x\|, \frac{x}{\|x\|}\right) = (x,\theta)$ to get

$$T\left(\left(X_{ij}/b_0\,(n/k)\right)^{\alpha}, i = 1,\ldots,n\right) = \left(\left(R_{i,k},\Theta_{i,k}\right);\ i = 1,\ldots,n\right)$$

and from (6.4.9) we get the estimator

$$S_n^0\,(\cdot) = \frac{\sum_{i=1}^{n} \epsilon_{\{R_{ik},\Theta_{ik}\}}\left(\left(W_{n,k},\infty\right] \times \cdot\right)}{\sum_{i=1}^{n} \epsilon_{\{R_{ik},\Theta_{ik}\}}\left(\left(W_{n,k},\infty\right]\right)} \Rightarrow S^0\,(\cdot)\,. \tag{6.4.17}$$

The interpretation of (6.4.17) is to chose Θ_{ik} whose radius is greater than $W_{n,k}$. The hidden angular measure is in this case defined on the interval $[0,\pi/2]$. For instance if the most of the mass density concentrates around $\pi/4$ then the dependency between extreme movements of the marginals is stronger, while the most of the mass around 0 and $\pi/2$ shows a tendency towards independence.

The number of order statistics k, the tail indeces of the marginals α_j and the parameter $b\,(n/k)$ can be estimated as in Proposition 6.4.14. We discuss the some statistical aspects in connection with the estimators, as asymptotic normality and other forms of regular variation.

6.4.3. Asymptotic normality of the estimators and its relation with second order of regular variation. This subsection shows asymptotic normality of the tail empirical measure and estimators based on second order regular variation, which plays an important role in the classical framework as in hidden regular variation.

Definition 6.4.16. (Second Order Regular variation ($2\mathcal{R}$)) A distribution function is second order regularly varying, with parameters $-\alpha$ and ρ ($1-F \in 2\mathcal{R}_{-\alpha,\rho}$) if there exists a function $A(t) \to 0$, $t \to \infty$ such that $|A(t)| \in \mathcal{R}_\rho$ and

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\alpha}}{A(t)} =: H(x) \tag{6.4.18}$$

holds, where $H(x) := cx^{-\alpha}\int_1^x u^{\rho-1}du$ for $x > 0$.

In a bivariate framework the equation (6.4.18) can be written similarly in terms of the tail measure as

$$\lim_{t\to\infty} \frac{t\mathbb{P}\left(\frac{X}{b_1(t)}, \frac{Y}{b_2(t)} \in ([0,x]\times[0,y])^c\right) - v_\alpha\left(([0,x]\times[0,y])^c\right)}{A(b_1(t), b_2(t))} =: H(x,y), \tag{6.4.19}$$

where v_α is defined in the cone $\mathbb{C}$.

This second order characterization of regular varying functions has shown to be very useful in the analysis of asymptotic normality of extreme value statistics, see for instance (Haan and Resnick (1998); De Haan and Ferreira (2006); Resnick (2006)).

The next proposition resumes the asymptotic behaviour of the tail empirical measure.

Proposition 6.4.17. *(Asymptotic normality of the tail empirical measure) Let* $\mathbf{X}$ *be a random variable with distribution* F *whose tail is regularly varying. Then, for* $k \to \infty$ *and* $k/n \to 0$*, we have*

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{b(n/k)^{-1}X_i}\left(x^{-\alpha}, \infty\right] - \frac{n}{k}\left(1 - F\left(b(n/k)\,x^{-1/\alpha}\right)\right)\right) \Rightarrow W(x)$$

in $D((0,\infty])$*, where* $W(x)$ *is a standard Brownian motion and* $D((0,\infty])$ *is the space of real-valued, right continuos functions with finite limits existing on* $(0,\infty)$*.*

Proof. There exist different proofs for this standard result, see for example Resnick (2006, page 292.) □

A similar result can be derived if we include the second order condition of regular variation.

Proposition 6.4.18. *Let* $1-F \in \mathcal{R}_\alpha$*, We say that* $1-F \in 2\mathcal{R}_{-\alpha,\rho}$ *if for* $\psi = 2|\rho|/(\alpha + 2|\rho|)$ *there exists a function* $P \in \mathcal{R}_\psi$ *such that* $P(t) \to \infty$ *as* $t \to \infty$ *and there exist a function* $\varrho(x) := \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\alpha}}{A(t)}$ *such that for* $k = [P(t)]$ *we have that for each* $x \geq 0$*.*

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{b(n/k)^{-1}X_i}(x,\infty] - x^{-\alpha}\right) \Rightarrow W\left(x^{-\alpha}\right) + \varrho(x).$$

Untill now, we discussed the relation between second order regular variation and hidden regular variation. In the case of asymptotic independence we have that equation (6.4.19) is reduced to $t\mathbb{P}\left(\frac{X}{b_1(t)}, \frac{Y}{b_2(t)} \in ([0,x]\times[0,y])^c\right)/A(b_1(t), b_2(t))$.

Supposing the limit exists and it is non-zero, we know that for a relatively compact rectangle $\Lambda \in \mathbb{C}_0$ the limit

$$\frac{t\mathbb{P}\left(\frac{X}{b_1(t)}, \frac{Y}{b_2(t)} \in \Lambda\right) - v_0(\Lambda)}{A(b_1(t), b_2(t))} := H_0(x,y)$$

holds in $M_+(\mathbb{C}_0)$.

Furthermore, the function $A(b_1(t), b_2(t))$ plays the most important role in the determination of the hidden tail index. Note that $t/A(b_1(t), b_2(t)) \in \mathcal{R}_{1-\rho/\alpha}$ is asymptotically equivalent to a strictly increasing function ϕ whose inverse function $\phi^{\leftarrow} \in \mathcal{R}_{\alpha/\alpha-\rho}$.

Let us define $b_0(t) := b \circ \phi^{\leftarrow}(t) \in \mathcal{R}$, since hidden regular variation can only hold if $b(t)/b_0(t) \to \infty$. Comparing with our assumption

$$\begin{aligned}
\lim_{t\to\infty} \frac{b(t)}{b_0(t)} &= \lim_{t\to\infty} \frac{b(t)}{b \circ \phi^{\leftarrow}(t)} = \lim_{t\to\infty} \frac{t}{\phi^{\leftarrow}(t)} \\
&= \lim_{t\to\infty} \frac{\phi(t)}{t} = \lim_{t\to\infty} \frac{t/A(b(t))}{t} \\
&= \lim_{t\to\infty} \frac{1}{A(b(t))} \to \infty,
\end{aligned}$$

shows that our election for $b_0(t)$ is correct.

Moreover, $b_0(t) \in \mathcal{R}_{1/\alpha_0}$ from Definition 6.4.2.1 and in the standard case $\alpha = 1$, therefore the tail index of the hidden regular measure is $\alpha_0 = 1 - \rho$.

In conclusion, we observe that under the classical definition of regular variation plus the condition of asymptotic independence, i.e,. $v_\alpha(\mathbb{C}) = 0$ in $M_+(\mathbb{C})$ and the second order regular variation are sufficient to obtain hidden regular variation.

On the contrary to the classical multivariate regular variation in $\mathbb{C}$, where one natural form on testing independence between marginals is through consistency and the asymptotic normality of the empirical measure. In the case of asymptotic independence

$$\sqrt{k}\left(\mu(x_1, x_2) - (x_1 + x_2)\right) \to 0,$$

see for instance Resnick (2006); De Haan and Ferreira (2006).

Thus, $\hat{\mu}(x, y) = \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left\{\frac{X_1}{b_1(n/k)}, \frac{X_2}{b_2(n/k)}\right\}}\left(\left[0,\, x_1^{-\alpha_1}\right] \times \left[0,\, x_2^{-\alpha_2}\right]\right)^c$ cannot be used to construct a test statistic. To avoid the asymptotic variance to vanish Hüsler and Li (2008) propose to divide the sample into two sub-samples with equal sub-sample size. Then, they use the second sub-sample to estimate the tail quantiles for each marginal, and define a new estimator of $\hat{\mu}(x, y)$ using the first sub-sample. Since the two sub-samples are independent. the asymptotic variance does not vanish. We use this idea in our framework to derive a test of asymptotic independence. This idea will be resumed in next theorem.

THEOREM 6.4.19. *let* $\mathbf{X} = \{X_1, X_2\} = \{(X_{1,1}, X_{2,1}), \ldots, (X_{1,n}, X_{2,n}), \ldots, (X_{1,m}, X_{2,m})\}$ *be a iid bivariate random vector with distribution function* F *and copula function* $C(x_1, x_2)$ *. Then, there exist* $0 < \varepsilon < 1$ *and* $0 < \epsilon < 1$ *such that* $x_1 - C(x_1, x_2) \geq \epsilon x_1$ *and* $x_2 - C(x_1, x_2) \geq \epsilon x_2$ *for* $x_1, x_2 \in (0, \varepsilon]$*. Furthemore, let* $k = o\left(n^{2\alpha/(1+2\alpha)}\right)$ *for some* $\alpha > 0$ *and* $\theta = m/n$*.*

Now, define the new estimator of μ_1*by*

$$\hat{\mu}^{n,m}(x_1, x_2) = \frac{m}{k}\frac{1}{n}\sum_{i=1}^{n} \epsilon_{\left\{\frac{X_1}{b_1^m(n/k)}, \frac{X_2}{b_2^m(n/k)}\right\}}\left(\left[0,\, x_1^{-\alpha_1}\right] \times \left[0,\, x_2^{-\alpha_2}\right]\right)^c.$$

Then , under asymptotic independence

$$\sup_{x_1, x_2} \sqrt{k}\left|\hat{\mu}^{n,m}(x_1, x_2) - \left(x_1^{-\alpha_1} + x_2^{-\alpha_2}\right)\right| \to \sup_{x_1, x_2} \left|W_1((1+\theta)x_1) + W_2((1+\theta)x_2)\right|$$

as $n \to \infty$*, where* W_1 *and* W_2 *are two independent Brownian motions.*

As we can see this theorem is similar to the Kolmorov-Smirnov test. This can be used to construct a test of asymptotic independence for real data. In fact, we use the quantiles of a simulate 200.000 Brownian motions for $\sup_{x_1,x_2} |W_1((1+\theta)x_1) + W_2((1+\theta)x_2)|$ with $\theta = 1$. If our estimated measure $\hat{\mu}_1^{n,m}(x_1,x_2) - \left(x_1^{-\alpha_1} + x_2^{-\alpha_2}\right)$ is smaller than the $(1-\alpha)$-th quantile of our simulations, then we have no reason to reject the hipothesis of asymptotic independence with confidence interval $(1-\alpha)$. Otherwise we may reject it.

6.4.4. Simulations. A summary of estimations of the hidden regular tail index α_0 obtained through the Hill's estimator from simulated data are resumed in Tables 6.B.1 and 6.B.2. We calculate the hidden regular tail index of a bivariate normal and a bivariate logistic dependence for different levels of dependence, threshold levels and several samples size.

For each distribution we generate 500 samples, in Tables 6.B.1 and 6.B.2, besides the mean, the standard deviation and the root mean squared (RMSE) of the estimates. The true values of the tail index are $\alpha_0 = 2/(1+\rho)$ for the bivariate normal distribution, where ρ is the correlation, and for the logistic distribution $\alpha_0 = 1$ for all $\beta < 1$ (see Ledford and Tawn (1996))

For the bivariate normal dependence structure, when the correlation function is $\rho = 0.1$ or $\rho = 0.9$, and the threshold level and the sample sizes increase, the dependence is surprisingly better estimated in terms of RMSE. Furthermore, when we increase the threshold and the sample sizes the bias are reduced and the resulting tail index α_0 is more consistent with the true values.

In the case of the bivariate logistic distribution, we observe that while the exponent $\beta \leq 0.7$ the results coincide with the empirical findings. However, for $\beta = 0.9$ the resulting values for α_0 are more consistent with asymptotic independence. Observe, that the best estimation for this case is when the sample size is large and the threshold is low. This is an extreme case, but it illustrates the difficulty when having weak asymptotic dependence.

6.5. "Let the tail go with the hidden": contagion, linkages between Brazil and Russia

Globalization process has facilitated the transmission channels of financial concerns beyond the physical borders. It is a central issue in asset allocation and risk management is whether financial markets become more interdependent during financial crises. This issue acquired dramatic importance during the last crises, as the Russian default in 1998, the devaluation of the Brazilian real in 1999, or the most recent Subprime Crisis (2007- 2009).

Within the crises which generated the most contagion are the Russian and U.S. subprime crises, which both began in credit markets and spread to stock markets. In the Russian crisis, Russia owed almost $20 billion in foreign debt which had to be paid back in 1998. The East Asian financial crisis in 1997 which affected the export revenues of Russia severely and the subsequent fall of global oil prices increased a financial burden on budget and led to a reduction in Russian foreign reserves. This fragile situation forced the Russian government in the middle of 1998 to halt the repayment of its foreign debts. The effects of such a decision was a growing trend in capital outflows from the economy which by IMF prescription on increasing interest rates meant an extra burden on the fragile banking system of Russia. In August 1998, the

Russian government announced its inability to control the value of the Ruble. This currency lost its value overnight by more than 100 per cent, from 7 Rubles per dollar to about 17 Rubles per dollar. In the month after the 1998 devaluation of the Russian Ruble, the Brazilian stock market fell by over 50 per cent. All these events happened during a time where there was not a considerable trade and financial linkage between Russia and Brazil. The question is whether such a reaction in Brazil can be interpreted as a contagion from Russia. Was Brazilian crisis caused by the Russia default? Were the other Latin-American countries also affected by those shocks in Russia and Brazil?.

On the other hand, the meltdown in the US subprime real-estate market has led until now to a global loss of more 7.7 trillion dollars in stock market value since October 2007. The crisis, which has spread beyond US shores to banks and other sectors worldwide, is one of the most vicious in financial history. The losses are worse than any in the past few decades, including Wall Street's Black Monday of 1987, the 1999 Brazilian real currency crisis and the collapse of hedge fund Long Term Capital Management (LTCM) in 1998. The losses are also greater than those suffered after the September 11, 2001, terror attacks, the Asian financial crisis starting in 1997, Argentina's default on its debt in 2001 and the 1994 Mexican peso crisis. It will take months or even years before Wall Street gets a handle on true cost of the US subprime meltdown and the attendant global credit crunch.

In the cases of Russia and Brazil in the recent crisis, their stocks were close to record peaks, its foreign reserves were the envy of the world and some analysts were even describing it as a heaven for investments. Today, largely due to the bursting of the commodity bubble that had underpinned many emerging market economies around the world, the equity market in Russia has fallen a dramatic 70 per cent since May 2008 amid the sharp sell-off in commodity related stocks, while for many of the Latin American countries, such as Brazil, the stock market has dropped 56 per cent this year.

In this section, our effort is to investigate the common wisdom about the existence of contagion by a new approach and expanding the common knowledge about this important event. Of course, the identification of shocks triggering a crisis is just one dimension to understand financial crises. A second and arguably more important dimension, is to identify the transmission mechanisms that propagate shocks from the source country across national borders and across financial markets. However, we wish to consider only the first point.

In particular we employ the estimators defined above to two periods of crises in which contagion effects link markets across national borders and asset classes could be found. The crises considered are Russia Flu in the second half of 1998, Brazil in early 1999, and the most recent crisis, as for example, those suffered after the September 11, 2001, terror attacks and U.S. subprime mortgage and credit crisis from 2007 until today.

We use daily stock and bond returns on these two emerging markets from 1995 to 2008 to seek for extreme spillovers. In other words, we want to calculate a linkage measure which indicates whether markets move together in turbulent periods or not. Therefore we would be well advised to turn to a measure which is not conditioned on a particular multivariate distribution and which directly reflects the probabilities and associated crisis levels.

6.5.1. A first definition of contagion. One of the most interesting aspects of the existing literature on contagion is inconclusive agreement on the definition of contagion. Calvo and Reinhart (1996) emphasized on the "true contagion" as "non-fundamental based" transmission mechanisms of shocks, while on the basis of their approach, transmission of shocks through economic fundamental links would be defined as "spillovers". Fratzscher (2002) also emphasizes that contagion is a transmission of crisis which is not caused by the affected country's fundamental. In contrast to these two ideas, Moser (2003) mentions that true contagion is the only response to fundamental-based contagion. The other area of confusing is separation of "contagion" and "interdependence". Interdependence implies that both markets collapse because of a influence of a common factor (Forbes (2002),Corsetti et al. (2005)). Forbes (2002) define contagion as significant increase of cross-market co-movement after shock. The lack of "significant" increase and only continued high level of market correlation distinguish contagion and interdependence in their approach.

Contagion is defined as a significant increase in cross-market linkages measured by tail dependence after a shock to a specific market of a country. Furthermore, if the structure of transmission mechanisms is found to be common across different times of crises, this would suggest that all crises are indeed alike regardless of the nature of the initial shock and the economic and institutional environments of the affected country. Alternatively, if the propagation mechanisms vary across crises, perhaps as a result of the development of new strains of contagion, this would suggest that crises are indeed unique at least across their source and their transmission mechanics.

6.5.2. Literature review of contagion between Russia and Brazil. One of the main studies which explicitly examines the possible contagion from Russia to Brazil carried out by Baig and Goldfajn (1999). They analyse the key players and timing of events of this crisis. The main reason which they offer for the Brazilian crisis is the foreign investors panicky behaviour after Russian default. The reaction of these investors in addition to local investors in Brazil destabilized the currency market of this country as well. Furthermore, they calculate the heteroscedasticity-adjusted correlation coefficients among rates of return on Brady bonds and find a significant increase in these rates after Russian crisis. In their opinion, the off-shore Brady market was the most possible channel of claimed contagion from Russia to Brazil. However, by using monthly data on aggregated data (total flows), they were not able to support the existence of contagion from Russia to Brazil.

Krugman (1999) also refers to panic among the hedge funds after Russian crisis. He also raises this question "What does Brazil have to do with Russia?". In response to this question he mentions that both countries had the same international investors. When these hedge funds lost their money in Russia, other lenders became more sensitive to other emerging economies and pulled out their money from Brazil. The main reason behind the contagion, Krugman (1999) says, was psychological.

Sull (2006) has also examined the possible contagion from Russia to Brazil. He analyzed the adjusted correlation between the Russian and Brazilian daily stock market indexes over the period of 1997-1999. After adjusting for the increase in volatility, he shows that correlation for Russia and Brazil stock market remained at a steady level and no significant increase realized in August 1998. Therefore, his results cannot support the common wisdom of contagion from Russia to Brazil.

Within a nonparametric framework for tail dependence in the constant conditional correlation GARCH case, Herrera et al. (2008)has concentrated principally on the contagion between Russia and Brazil in the late 1998, and the potential contagion from Brazil to other countries of Latin America. They find a true contagion from Russia to Brazil through an empirical measure of extreme dependence to understand the joint extremal behaviour of multivariate time series.

6.5.3. Data and stylized facts. We will consider the same models proposed in Section 6.4.2 for the two countries under study. We model the equity and bond markets in Brazil and Russia. The series for the Brazil equity returns is the Bovespa Index, while for the Russian equity returns is the RTSI index. In the case of the bond markets we have used JP Morgan EMBI Global Index. Our study employs daily data on the mentioned indices during the period 1.09.1995 until 10.10.2008 in order to evaluate the probability of contagion between Russia and Brazil. All return series were generated using the continuous compounding formula $X_t = \ln(P_t/P_{t-1})$, where P_t represents the price series at time t.

We begin our investigation by trying to establish the two sample periods to study. We want to be able to define the two crisis episode in order to compare results and analyze the factors behind the movements. Given the numerous shocks that financial markets have faced in the last teen years, isolating different periods of crises is somewhat difficult.

We choose arbitrary as the end of the first period sample 31.01.2001 to have nearly the same number of observations in the two samples. Note that the first period also includes the Asian crises of 1997 and the Dot.com crash of 2000, while the second period also includes the terror attacks from September 11, 2001.

During this study we concentrate only on the negative cone for each pair of possible combinations among the different countries and assets. i.e., the results presented pertain in all cases to the lower tails of the return distributions, these are the extreme losses. The scatter plots for the negative cones are displayed in the Figures 6.5.1 and 6.5.2 for the first and second period respectively.

In the first period one can observe that for the 0.995-th empirical quantile of each bivariate combination the most extreme events occurred during the Russian Crises with some exceptions during the Asian crises during 1997. They are displayed with black colors. Furthermore, we find that these extremes tend to cluster, which is indicated with arrows between consecutive days. However, it is not clear that the extreme events in the different asset classes tend to happen together.

In the second period the most extreme events took place during the actual Subprime Crises with some exceptions during the terror attacks in September 11, 2001. Like in the first period, the extremes events tend to cluster, while the dependence among these extreme events seems to decrease between stock and bond markets of different countries. A more detailled analysis will be given later.

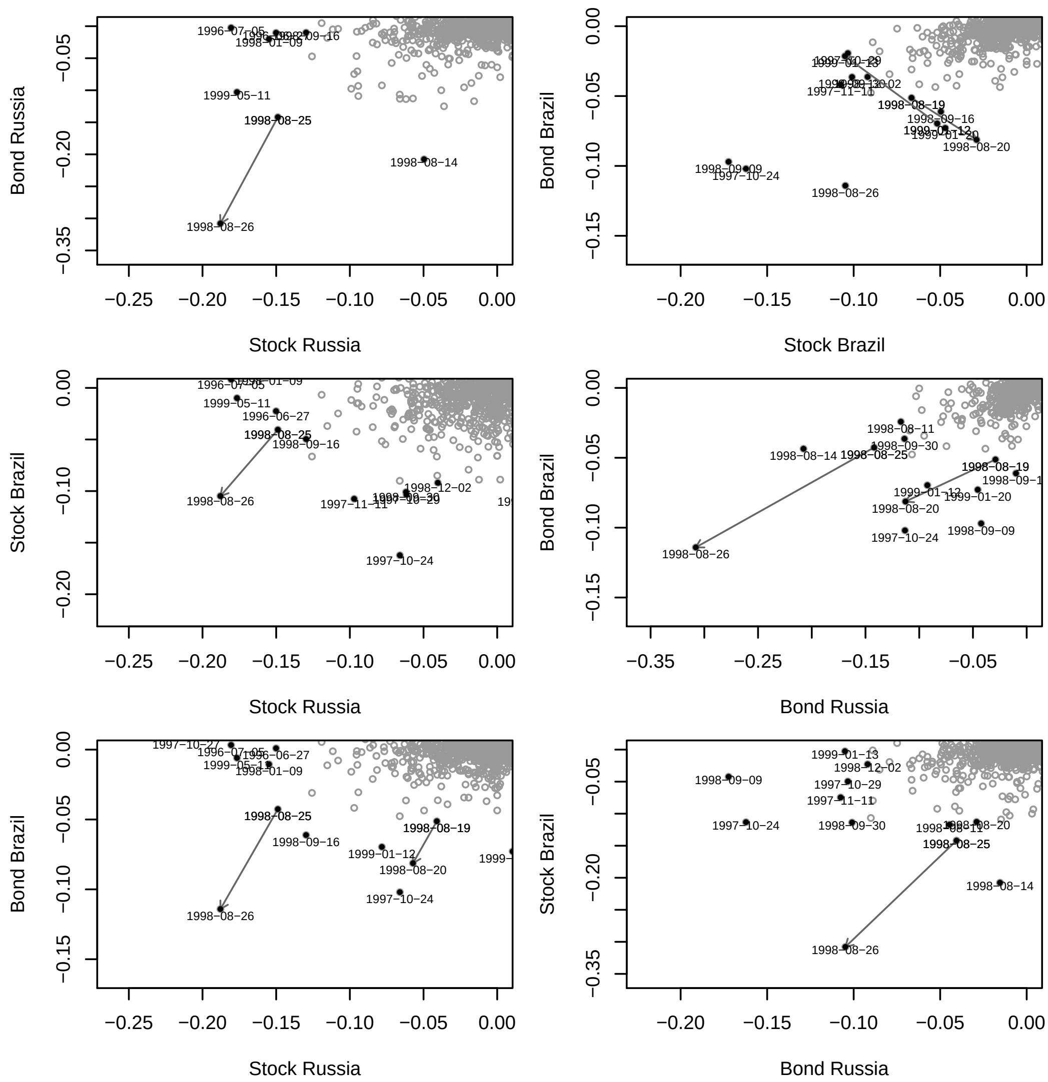

FIGURE 6.5.1. Scatter plots for the different pairs of asset combinations between Russia and Brazil for the fist period. The most extreme events (0.95-th empirical quantile) are displayed in black colours with their correponding date. The extreme events between consecutive days are indicated with arrows.

A summary of statistics for the data is given in Table 6.B.3. The Jarque-Bera test statistic provides clear evidence to reject the null hypothesis of normality for the unconditional distribution of the daily returns. The high value for excess kurtosis indicates that the distributions are characterized by leptokurtosis. Moreover, we compute the augmented Dickey-Fuller test for the null hypothesis that the returns have a unit root. This hypothesis is also rejected for the whole series.

The difference to Herrera et al. (2008), where an empirical measure for the tail dependence was applied directly to the marginal distributions based on asymptotic results of the tail behaviour of stochastic recurrence equations, an univariate GARCH filter is used for the marginal distributions, due to the fact that these models are only appropriate for iid random vectors.

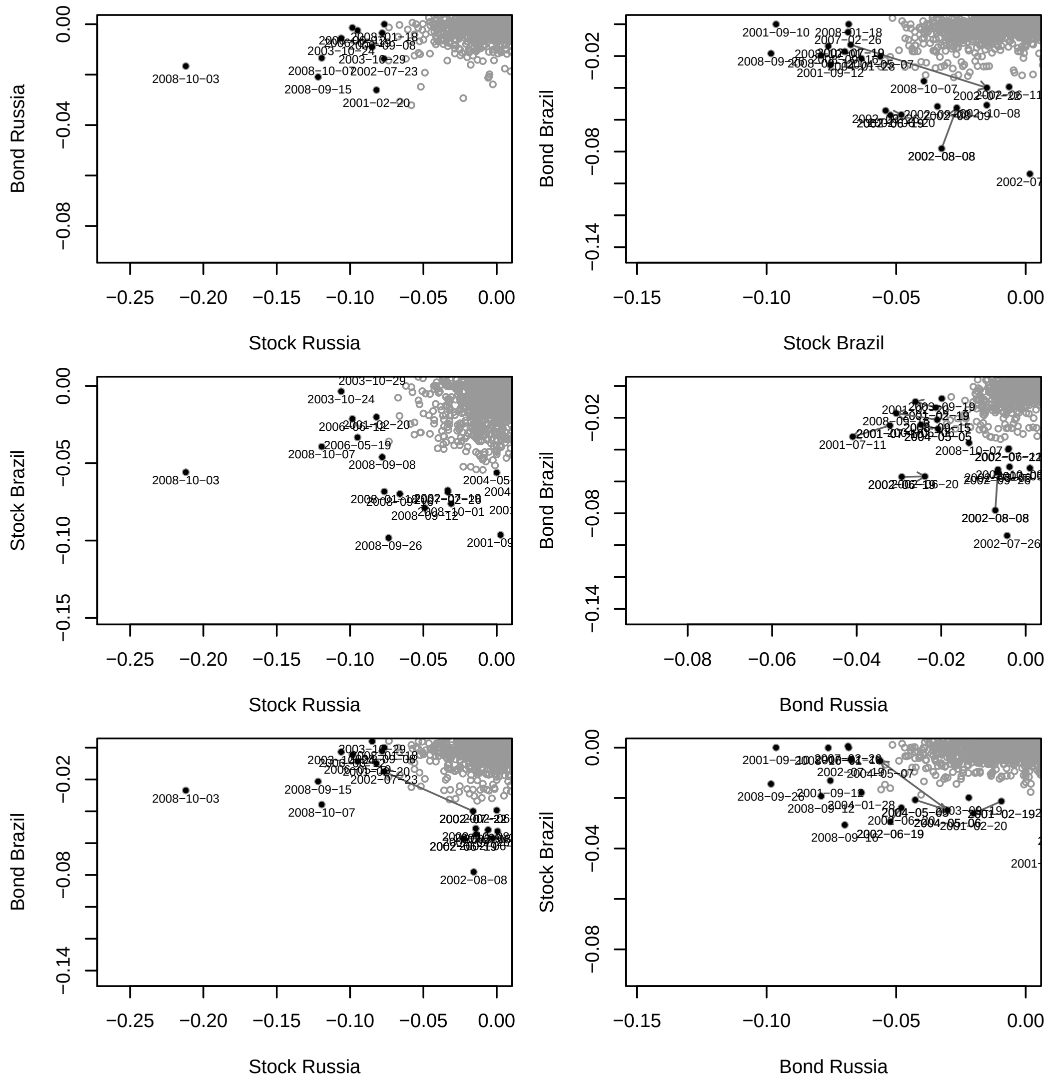

FIGURE 6.5.2. Scatter plots for the different pairs of asset combinations between Russia and Brazil for the second period. The most extreme events (0.95-th empirical quantile) are displayed in black colours with their correponding date. The extreme events between consecutive days are indicated with arrows.

Table 6.B.4 in the Appendix presents descriptive statistics for returns that are filtered for heteroskedasticity together with detail specifications of the volatility filters. We conclude from these results that the t-GARCH(1,1) may fully capture the time-varying volatility in the data, and therefore, our models are perfectly applicable to the data after the garch filter. The large values for the $\beta's$ parameters capture the high persistence in volatility. The values of the parameters α are not too high, which means that ARCH effects are weak.

6.5.4. Extreme dependence structure in bond and equity markets during the Russian crisis. In a first approximation and for the sake of completeness, we estimate the conditional correlation for pairs of assets under study assuming, that these pair follow the linear factor model in equation (6.4.2) for $\lambda_i = 1$, $u \sim N(0,1)$ and conditioning quantile set $t \in \Omega := [0.95, 0.99]$. The Figures 6.5.3 and 6.5.4 despict these results together with two special cases, when the linear factor model follows a Gaussian and a t-Studend model for the risk factor. For both figures the colour line indicates the conditioned market, while the dash lines give the same information when the conditioning is a Gaussian or a t-Studend distributed risk factor for both markets.

For the first period, the most pairs of combinations show significant correlation for lower quantiles whatever the conditioning variable may be. In contrast, conditioned on large fluctuations in higher quantiles, the correlation coefficient sinks significantly for the pairs Stock Brazil vs. Bond Brazil and Stock Brazil vs. Bond Russia. On the other hand, the conditional correlation remains stable for large negative returns in the bond markets.

Other interesting feature for the pairs Stock Russia vs. Bond Russia and Stock Russia vs. Bond Brazil is that their conditional correlations can be well approximated by a t-Studend risk factor. Furthermore, there is a strong and significant symmetry between the conditional correlations.

We have performed these estimations for the second period of the study as well. Contrary to the first period the results are more diverse and therefore difficult to interpret. The results which are based on the sample correlation are erratic and in the interval 0 to 0.5 for the most pairs, with some estimations even yielding to negative conditional correlation.

Thus, in the second period, the conditional correlations in the simple linear factor model do not appear to be very useful tools for examining the possible changes in the dependence structure between two assets.

We stress that it would be erroneous to conclude, from the estimations of the conditional correlations during the first period, a genuine increase in the dependence only from the interplay between the conditioning and the dependence structure of the factor model as we have shown analytically in Subsection 6.4.1. For example, the conditional correlations can increase conditioning on larger fluctuations without needing any variation of the unconditional correlation coefficient.

As consequence of the deficiencies of linear factor models with a conditional correlation framework to capture the extreme dependence, we omitted additional analysis in this methodology for characterizing possible changes in the dependence structure.

Following, we concentrate on the models proposed in Section 6.4.2 to the bond and stock market of these two countries. The first important step is to distinguish between asymptotically dependent and asymptotically independent variables to quantify the degree of dependence for the appropriate estimation of regions going to the infinity. In each of the following subsections sets of estimations and test statistics will be presented in graphical form to facilitate the interpretation.

In the Figures 6.B.1 and 6.B.2 we document the estimation of standardized tail index α_0 of hidden regular variation together with the starica plot for the choice of the number of order statistics to estimate the tail indices, the distribution of spectral measure $S(\cdot)$ and the hidden spectral measure $S_0(\cdot)$ for the first and second period of investigation respectively.

The estimation of the tail index before and after volatility filtering in the univariate case are resumed is the Table 6.B.5 together with the estimation of the tail indices for the hidden measure among international stock and bond market returns. The estimations suggest

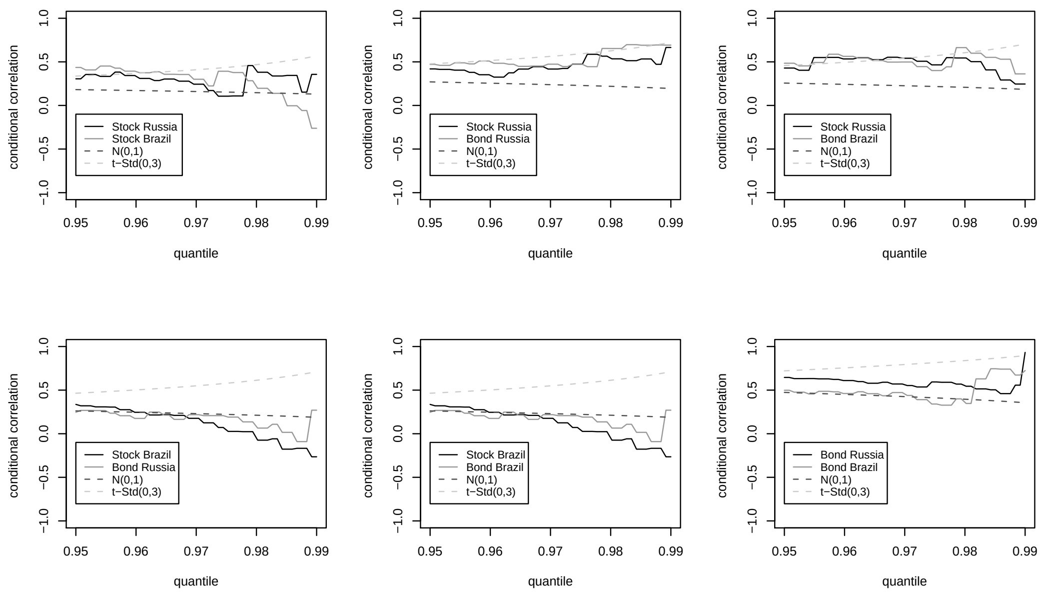

FIGURE 6.5.3. Conditional correlation estimate for each pair of combinations in a linear factor model in the first period of study. The two dashed lines represent the borders within which these conditional correlation coefficients could be represented by a linear factor model with Gaussian or t-Studend (with 3 degree of freedom) distributed risk factor.

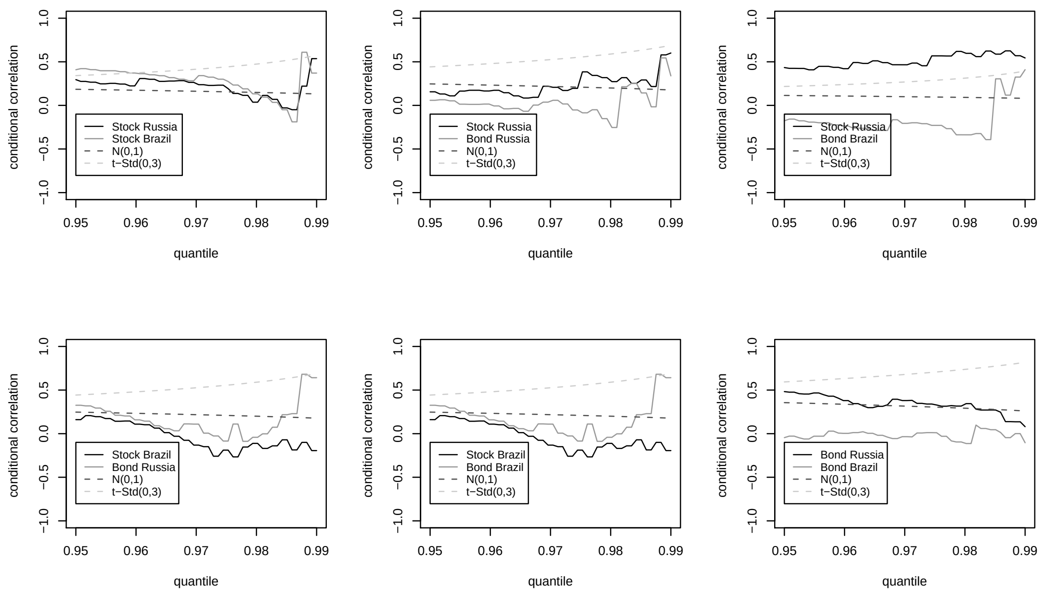

FIGURE 6.5.4. Conditional correlation estimate for each pair of combinations in a linear factor model in the second period of the study. The two dashed lines represent the borders within which these conditional correlation coefficients could be represented by a linear factor model with Gaussian or t-Studend (with 3 degree of freedom) distributed risk factor.

that the tail indices are significantly reduced when stock and bond returns are filtered for heteroskedasticity.

Notice that in the most of the cases α_0 is significantly more than 1 for the heteroskedasticity filtered returns. However, the confidence interval at the 95% is toowide to consider any pair of returns asymptotic independent with the exception of the extreme dependence between the two stock markets returns. A first interpretation of this result implies that there is significant dependence between large values of the paired series but that the largest values do not occur concurrently. Since most of the pairs have some degree of asymptotically independent, multivariate extreme value models that assume asymptotic dependence among the different market returns are likely to overestimate the joint risk.

In Figure 6.5.5 we assess how important the spillovers between the marginals of different markets and class of assets are, assuming some degree of asymptotic independence under the max-norm for different quantile levels.

Three main conclusions emerge from the results. First, we find extreme dependence between equities and bonds for the same country, i.e., intra country dependence, but this disappears rapidly to higher quantiles. That does not indicate contagion but perhaps flight to quality, because in the case of "flight to quality", dependence between stocks and bonds strongly decrease in fragile stock markets since this constitutes a movement of the asset classes in opposite directions. In the case of Brazil we observe that this extreme dependence is low.

Second, there appears to be a substantially larger extreme dependence between bonds than what we saw in the stock market case cross country. For instance, the spillover probability of experiencing a crash (at the 97.5-quantile) in the Brazilian bond market, given the same crash in the Russian bond market (at the 97.5-quantile) was circa 0.40. This means that in the case of experiencing two extreme movements in the bond market of Brazil, it would be reasonable to expect a similar behaviour in the Russian bond market. However, this probability slowly vanishes to higher quantiles as the rest of spillover probabilities. In the case of the stock markets cross country the possibility of the appearance of joint extreme movement is the lowest. In fact, they show the major degree of asymptotic independence.

Third, similar results can be observed by taking into account the probability of financial crisis among assets of different classes of different countries. Thus, the strength of these linkages are less important during the crises. As has indicated in literature (Forbes (2002)), there are no direct trade links between Russia and Brazil and the two countries do not export similar goods and have no strategic competition in their markets.

6.5.5. Extreme dependence structure in bond and equity markets after the Russian crisis until the subprime crises. In Table 6.B.5 we observe that most of tail indices of the hidden regular measure for the different combinations of assets and countries are much lower than the corresponding tail indices in the first period. While only two pairs estimates are significantly larger than in the first period of study, which means asymptotic independence has augmented in these cases. For instance, the asymptotic independence between the pairs Stock Russia vs. Bond Brazil and Bond Russia vs. Bond Brazil had been strength considerably.

In the case of intra-dependence in each country, the spillovers probability remains similar to the first period. In relation to the extreme dependence among assets of the same class in different countries, the probability of spillovers between the stock markets has increased slightly, while the probability of spillovers between the stock markets decreased.

These empirical findings have a direct impact on practices in the finance industry. For example, the fact that Stock Russia vs. Bond Brazil markets are asymptotically independent, while Stock Brazil vs. Bond Brazil are in the limit to be asymptotically dependent, means tail diversification and a reduction of portfolio extreme risk can better be achieved by holding Russian stocks and Brazilian bonds. However, the scope for tail diversification may be decreasing as our findings indicate that the asymptotic dependence cases, i.e., α_0 close to 1, have increased through the time.

These results reinforce the previous conjecture about the Pearson correlation as a poor measure for tail dependence. In the second period of the study the empirical results show that financial crises are not too different for these countries, as all linkages are statistically important across all crises. However, the strength of these linkages does vary across the two crises.

6.6. Conclusions

In this chapter we investigated among other things the consequences of the use of conditional correlation as measure of change of dependence. The results provide a quantitative proof of how the tail distribution functions behaviour can succeed or fail during a market decline as measure of extreme risk. The results given in the Subsection 6.4.1 are only simple examples but they make it painfully clear how misleading conditional correlation can be. In this respect, as recently stressed by Forbes (2002), many previous contagion studies which are based on linear factor models may be unreliable. We therefore need other methods to investigate whether for example more extreme movements in financial markets are indeed more highly depended than overall movements.

In a second contribution, it seems that the spillovers probabilities that we investigated in Subsection 6.4.2 could provide a useful measure of the extreme dependence among random variables. We use two nonparametric measures for extreme value dependence to characterise tail dependence or spillover probabilities in the asymptotic dependence and independence case.

We concentrated on the possible contagion case between Brazil and Russian markets and demonstrated how the tail risk measures may be assessed. These new tools have allowed us to document the asymptotic dependence and independence among stock and bond market returns of these countries. Notice, that the omission of asymptotic independence models in studies of contagion could led to mixed results and especially over-estimation could be possibly substantial in connection with measures of contagion. The empirical findings include a confirmation that extreme value dependence was much stronger in the first period of the study.

Ongoing research in these kind of measures should provide more interesting tools for risk management purposes since it gives directly the probability that an asset suffers a large loss assuming that a large loss occurred for another asset.

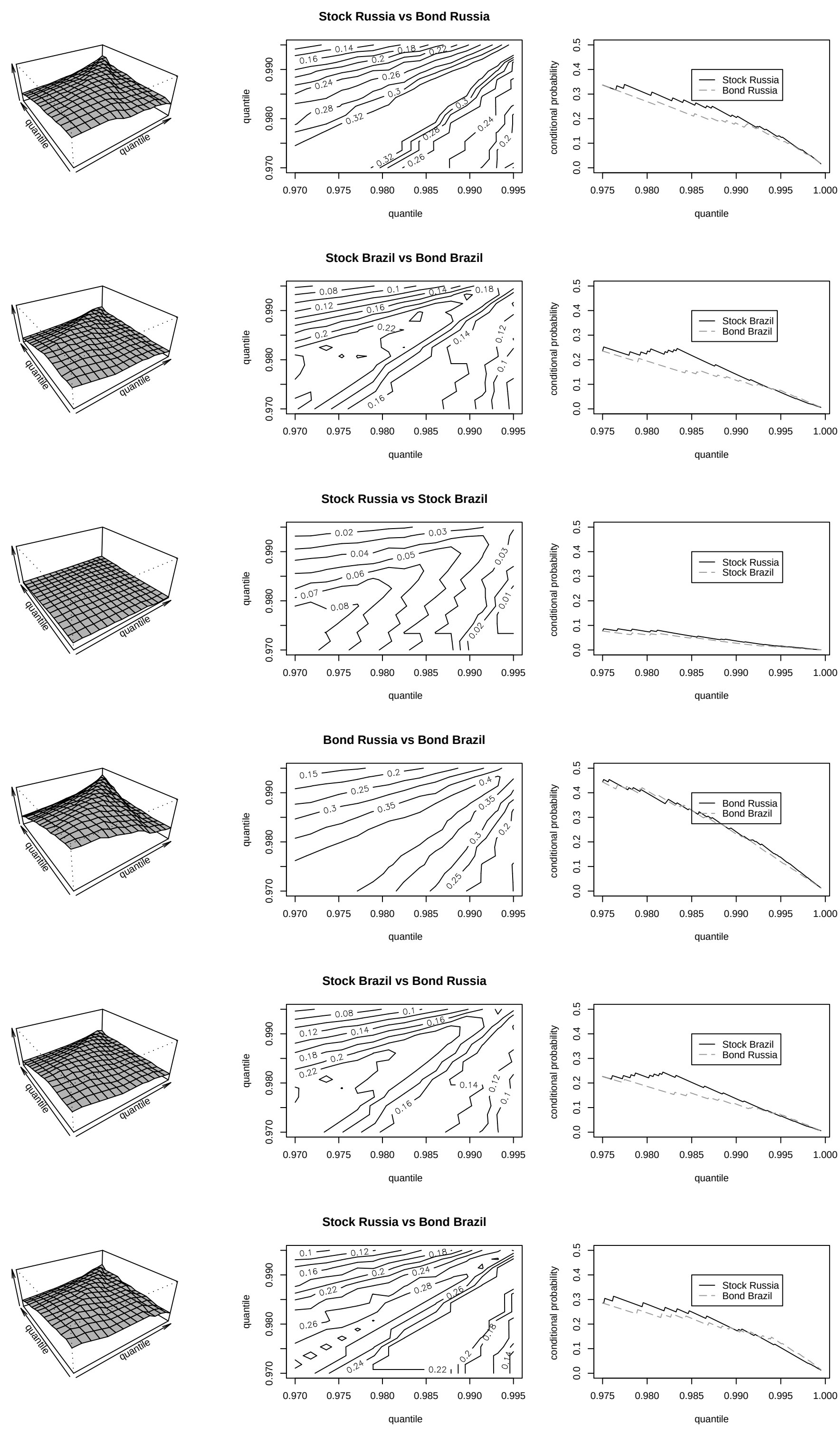

FIGURE 6.5.5. Conditional spillover probabilities for the empirical analysis in the first period of the study. Rigth column: Perpspective plot for the conditional spillover probabilities among different asset markets on different quantiles. The axis indicates the conditioned market. Middle column: Contour plot for the conditional spillover probabilities among different asset markets on different quantiles. The axis indicates the conditioned market. Left column: Conditional plot where in this case the threshold is fixed in 0.975th-quantile for the conditionating asset market, while for the conditioned market the threshold is variable. The lines represent the results of the spillover probability for the conditioned asset market. From top to bottom: Bond Russia vs. Stock Russia, Bond Brazil vs. Stock Brazil, Stock Brazil vs. Stock Russia, Bond Brazil vs. Bond Russia, Bond Russia vs. Stock Brazil, Bond Brazil vs. Stock Russia.

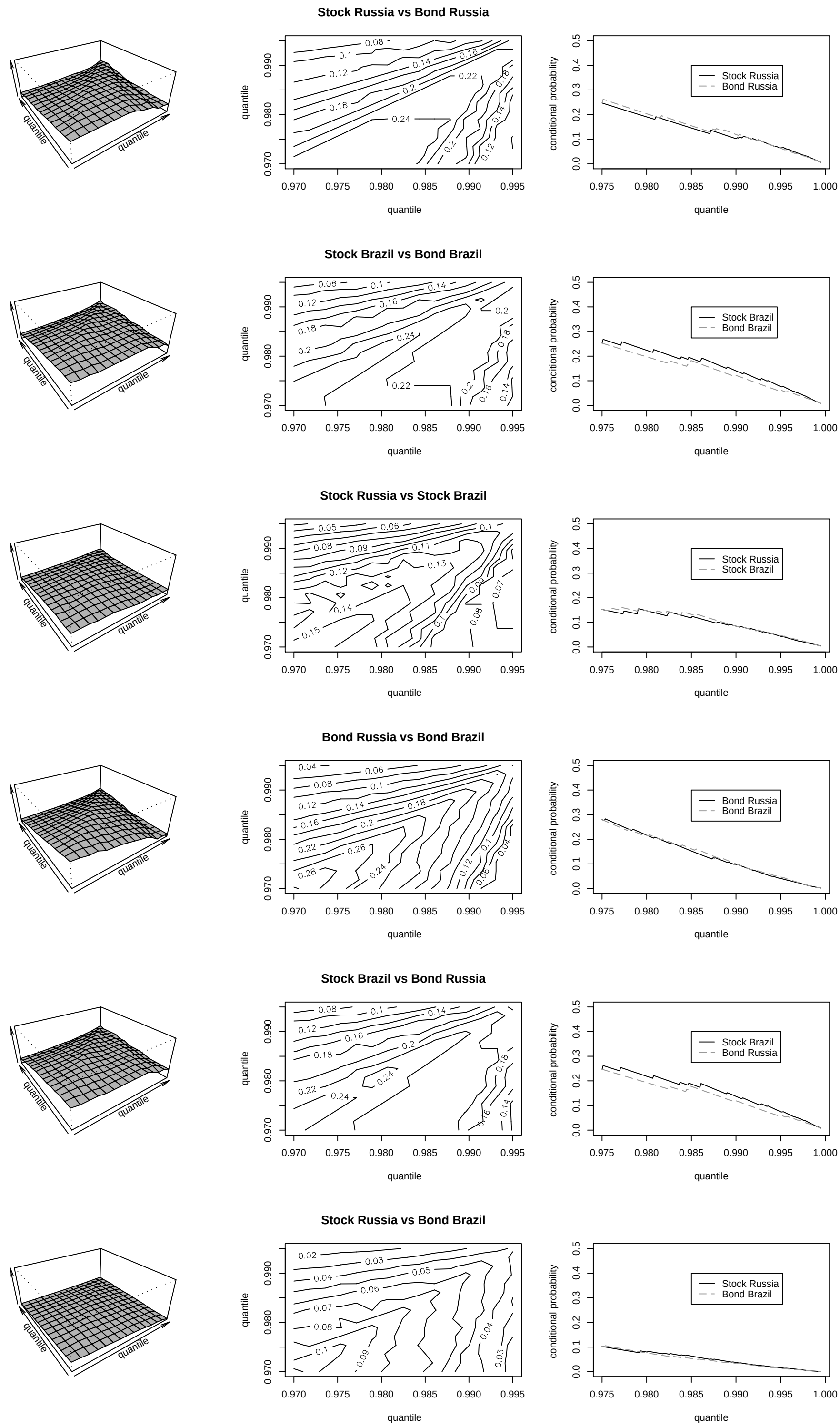

FIGURE 6.5.6. Conditional spillover probabilities for the empirical analysis in the second period of the study. Rigth column: Perpspective plot for the conditional spillover probabilities among different asset markets on different quantiles. The axis indicates the conditioned market. Middle column: Contour plot for the conditional spillover probabilities among different asset markets on different quantiles. The axis indicates the conditioned market. Left column: Conditional plot where in this case the threshold is fixed in 0.975th-quantile for the conditionating asset market, while for the conditioned market the threshold is variable. The lines represent the results of the spillover probability for the conditioned asset market. From top to bottom: Bond Russia vs. Stock Russia, Bond Brazil vs. Stock Brazil, Stock Brazil vs. Stock Russia, Bond Brazil vs. Bond Russia, Bond Russia vs. Stock Brazil, Bond Brazil vs. Stock Russia.

6.A. Demonstrations

Proof. Proposition 6.4.3: (1) Notice first that $a(x)^{-1} = \frac{d}{dx}\left(-\ln \overline{F}(x)\right) = \frac{f(x)}{\overline{F}(x)}$ for all $x < x_F$. Define the truncated k-th moment as $m_k(t) := \frac{\int_t^{\infty} x^2 f(x)dx}{\overline{F}(x)}$ of the random vector x.

An important characteristic of the derivates of $m_k(t)$ is that these can be represented in terms of $a(x)$ as

$$m_k'(t) = \frac{m_k(t) - t^k}{a(x)}.$$

Replacing these results in the conditional variance we get

$$\mathrm{Var}(w_t \mid w_t \in \Omega) := \Upsilon = m_2(t) - m_1(t)^2. \tag{6.A.1}$$

Taking the two first derivates of (6.A.1) we obtain the following two expansions:

$$\begin{aligned} \Upsilon' &= m_2'(t) - 2m_1'(t)\, m_1(t) \\ &= \frac{m_2(t) - t^2}{a(x)} - 2m_1(t)\frac{(m_1(t) - t)}{a(x)}, \end{aligned} \tag{6.A.2}$$

$$\begin{aligned} \Upsilon'' &= m_2''(t) - 2\left(m_1(t)\, m_2''(t) + m_2'(t)^2\right) \\ &= \frac{m_2(t) - t^2}{a'(x)} + \frac{m_2(t) - t^2 - 2t}{a(x)} - 2m_1(t)\frac{(m_1(t) - t)}{a'(x)} \\ &\quad -2m_1(t)\frac{(m_1(t) - t - 1)}{a(x)} - 2\frac{(m_1(t) - t)^2}{a(x)^2}. \end{aligned} \tag{6.A.3}$$

Now, by definition the function $a(x)$ have to be greater than zero, therefore, we can derive the next condition in (6.A.2)

$$m_2(t) - t^2 = 2m_1(t)\,(m_1(t) - t). \tag{6.A.4}$$

Including this result in the second derivate we obtain the following result

$$\frac{(m_1(t) - t)^2}{a(x)} + t + m_1(t) = 0. \tag{6.A.5}$$

Solving simultaneously equations (6.A.1), (6.A.4) and (6.A.5) we find $\Upsilon = a(x)^2$.

(1)$\Rightarrow$(2).

(3) It follows immediately by Karamatas Theorem 6.2.3. $\square$

Proof. Proposition 6.4.13: (1) Consistency of empirical measure given in the Definition (6.3.2) implies as $k \to \infty$ and $k/n \to \infty$, that $X_{n,k}/b(n/k) \to 1$. Notice that for a $\varepsilon > 0$, we have

$$\begin{aligned}\mathbb{P}\left(\left|\left(X_{n,k}/b\left(n/k\right)\right)-1\right|>\varepsilon\right) &= \mathbb{P}\left(X_{n,k}>\left(1+\varepsilon\right)b\left(n/k\right)\right)+\mathbb{P}\left(X_{n,k}<\left(1-\varepsilon\right)b\left(n/k\right)\right)\\ &\leq \mathbb{P}\left(k^{-1}\sum_{i=1}^{n}\epsilon_{\{X/b(n/k)\in(1+\varepsilon,\infty]\}}\geq 1\right)\\ &\quad +\mathbb{P}\left(k^{-1}\sum_{i=1}^{n}\epsilon_{\{X/b(n/k)\in(1-\varepsilon,\infty]\}}<1\right)\end{aligned}$$

and by Definition 6.3.2 we know that

$$k^{-1}\sum_{i=1}^{n}\epsilon_{\{X/b(n/k)\in(1+\varepsilon,\infty]\}}\to\left(1+\varepsilon\right)^{-\alpha}<1,$$

and

$$k^{-1}\sum_{i=1}^{n}\epsilon_{\{X/b(n/k)\in(1-\varepsilon,\infty]\}}\to\left(1+\varepsilon\right)^{\alpha}>1$$

and therefore the first proposition follows.

(3) The problem of the threshold selection can be affronted by means of the scaling property of the standardized empirical measure in equation (6.4.7). This idea was proposed by Starica (1999). He proposes to find a $\hat{k}$ such that $\hat{\mu}$ mimics the scaling property. In practice, we graphic $\frac{t\hat{\mu}_1\left(t\mathbb{S}^{d-1}\right)}{\hat{\mu}_1\left(\mathbb{S}^{d-1}\right)}$ for different values of $\hat{k}$ and choose one that seems to have the plot most close to the horizontal line at high 1.

(3) The Hill estimator is a classical method to find the tail index of a distribution function. Asymptotic normality and consistency were proven in Haan and Resnick (1998). $\square$

PROOF. Prposition 6.4.18: If $1-F\in 2\mathcal{R}_{-\alpha,\rho}$ and equation (6.4.18) holds then, necessarily $A\left(t\right)\in\mathcal{R}_{\rho}$, $b\left(t\right)\in\mathcal{R}_{1/\alpha}$ so that $A\left(b\left(t\right)\right)\in\mathcal{R}_{\rho/\alpha}$. We define the function

$$\chi\left(x\right)=\sqrt{x}/A\left(b\left(x\right)\right)\in\mathcal{R}_{(\alpha+2|\rho|)/2\alpha}$$

which implies that $\chi^{\leftarrow}\in\mathcal{R}_{2\alpha/(\alpha+2|\rho|)}$. Setting $P\left(t\right)=t/\chi^{\leftarrow}\left(\sqrt{t}\right)\in\mathcal{R}_{2|\rho|/(2\alpha+2|\rho|)}$, we have $P\left(t\right)/t\to 0$ as $t\to\infty$. Since we defined $P\left(t\right)\to\infty$, it is only possible if $|\rho|>0$ which is obvious. In another case note $P\left(t\right)\to\infty$ if and only if $t^2/\chi^{\leftarrow}\left(t\right)\to\infty$, which is equivalent to $\chi^2\left(t\right)/t\to\infty$, and to $1/A^2\left(b\left(t\right)\right)\to\infty$.

Observe that

$$\begin{aligned}\sqrt{k}\left(A\left(b\left(n/k\right)\right)\right) &= \sqrt{\left[P\left(n\right)\right]}\left(A\left(b\left(n/\right)\right)\right)\\ &= \sqrt{n}\frac{\left(A\left(b\left(n/\left[P\left(n\right)\right]\right)\right)\right)}{\sqrt{n/\left[P\left(n\right)\right]}}\\ &\sim \sqrt{n}\left(\chi\left(n/P\left(n\right)\right)\right)^{-1}\\ &= \sqrt{n}\left(\chi\left(\chi^{\leftarrow}\left(\sqrt{n}\right)\right)\right)^{-1}\end{aligned}$$

and as $n\to\infty$ we have $\sqrt{n}\left(\chi\left(\chi^{\leftarrow}\left(\sqrt{n}\right)\right)\right)^{-1}\to 1$.

Finally, we obtain the main result

$$\begin{aligned}
\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{b(n/k)^{-1}X_i}(x,\infty]-x^{-\alpha}\right) &= \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{b(n/k)^{-1}X_i}(x,\infty]-\frac{n}{k}\left(1-F\left(b\left(n/k\right)x\right)\right)\right) \\
&\quad +\sqrt{k}\left(\frac{n}{k}\left(1-F\left(b\left(n/k\right)x\right)\right)-x^{-\alpha}\right) \\
&= W\left(x^{-\alpha}\right)+o\left(1\right) \\
&\quad +\sqrt{k}A\left(b\left(n/k\right)\right)\left(\frac{\frac{n}{k}\left(1-F\left(b\left(n/k\right)x\right)\right)-x^{-\alpha}}{A\left(b\left(n/k\right)\right)}\right) \\
&= W\left(x^{-\alpha}\right)+\varrho\left(x\right)
\end{aligned}$$

$\square$

6.B. Tables and Figures

		$\rho = 0.1,\ \alpha_0 = 1.818$				$\rho = 0.4,\ \alpha_0 = 1.428$				$\rho = 0.7,\ \alpha_0 = 1.176$				$\rho = 0.9,\ \alpha_0 = 1.052$			
Size	q	Mean	Sd	RMSE	H_0	Mean	Sd	RMSE	H_0	Mean	Sd	RMSE	H_0	Mean	Sd	RMSE	H_0
500	0.03	2.000	0.491	0.526	0	1.601	0.406	0.441	0	1.322	0.277	0.312	0	1.205	0.198	0.251	0
1000	0.03	1.907	0.295	0.308	0	1.555	0.224	0.257	0	1.302	0.193	0.231	0	1.151	0.124	0.157	0
5000	0.03	1.861	0.142	0.148	0	1.506	0.113	0.137	0	1.256	0.081	0.114	0	1.112	0.050	0.078	3
10000	0.03	1.853	0.101	0.107	0	1.502	0.077	0.107	0	1.251	0.054	0.092	0	1.107	0.036	0.065	34
500	0.05	1.919	0.354	0.368	0	1.561	0.275	0.305	0	1.323	0.201	0.248	0	1.170	0.125	0.171	0
1000	0.05	1.889	0.238	0.248	0	1.540	0.196	0.225	0	1.277	0.130	0.165	0	1.140	0.086	0.122	0
5000	0.05	1.853	0.103	0.109	0	1.516	0.082	0.120	0	1.259	0.059	0.101	0	1.114	0.039	0.073	26
10000	0.05	1.857	0.076	0.085	0	1.506	0.058	0.097	0	1.254	0.038	0.087	0	1.109	0.026	0.062	421
500	0.1	1.878	0.231	0.238	0	1.561	0.189	0.230	0	1.306	0.136	0.187	0	1.147	0.084	0.127	0
1000	0.1	1.872	0.168	0.176	0	1.524	0.125	0.157	0	1.280	0.089	0.136	0	1.136	0.055	0.100	0
5000	0.1	1.859	0.076	0.086	0	1.519	0.057	0.107	0	1.266	0.039	0.097	34	1.116	0.024	0.068	486
10000	0.1	1.855	0.049	0.062	0	1.512	0.038	0.092	0	1.262	0.027	0.090	480	1.113	0.018	0.063	500

TABLE 6.B.1. Estimations of the hidden tail index α_0 for the bivariate normal distribution. The first column indicates the sample size. The second column describes the upper quantile of sample used to calculate the hidden tail index. The other columns represent the mean, the standard deviation and the root mean squared (RMSE) of the estimates in relation to the true values.

		$\beta = 0.1,\ \alpha_0 = 1$				$\beta = 0.4,\ \alpha_0 = 1$				$\beta = 0.7,\ \alpha_0 = 1$				$\beta = 0.9,\ \alpha_0 = 1$			
Size	q	Mean	Sd	RMSE	H_0	Mean	Sd	RMSE	H_0	Mean	Sd	RMSE	H_0	Mean	Sd	RMSE	H_0
500	0.03	1.097	0.074	0.122	0	1.112	0.142	0.181	52	1.152	0.231	0.276	101	1.370	0.353	0.511	201
1000	0.03	1.064	0.045	0.078	0	1.068	0.091	0.114	134	1.112	0.147	0.184	243	1.346	0.228	0.414	356
5000	0.03	1.020	0.017	0.026	52	1.025	0.040	0.047	257	1.064	0.060	0.088	345	1.294	0.102	0.311	500
10000	0.03	1.011	0.012	0.016	427	1.017	0.027	0.032	500	1.053	0.043	0.069	500	1.286	0.074	0.295	500
500	0.05	1.069	0.050	0.085	0	1.087	0.110	0.140	123	1.138	0.160	0.211	145	1.447	0.304	0.541	234
1000	0.05	1.045	0.031	0.055	0	1.053	0.066	0.085	278	1.113	0.115	0.161	295	1.394	0.182	0.434	420
5000	0.05	1.013	0.013	0.019	308	1.023	0.028	0.036	379	1.081	0.047	0.094	453	1.369	0.085	0.379	500
10000	0.05	1.008	0.009	0.012	500	1.019	0.021	0.028	500	1.081	0.034	0.088	500	1.356	0.058	0.361	500
500	0.1	1.045	0.032	0.055	250	1.067	0.066	0.094	245	1.163	0.118	0.201	345	1.501	0.189	0.535	420
1000	0.1	1.028	0.020	0.034	352	1.050	0.051	0.071	401	1.151	0.086	0.174	432	1.484	0.137	0.503	451
5000	0.1	1.009	0.009	0.013	500	1.028	0.021	0.035	500	1.132	0.036	0.137	500	1.471	0.060	0.475	500
10000	0.1	1.006	0.006	0.008	500	1.026	0.014	0.030	500	1.129	0.026	0.132	500	1.466	0.043	0.468	500

TABLE 6.B.2. Estimations of the hidden tail index α_0 for the bivariate logistic distribution. The first column indicates the sample size. The second column describes the upper quantile of sample used to calculate the hidden tail index. The other columns represent the mean, the standard deviation and the root mean squared (RMSE) of the estimates in relation to the true values.

Asset markets	mean	sd	min	max	skewness	kurtosis	Box.test	J.B-test	ADF	Engle (10)
Stock Brazil	624×10^{-6}	0.027	-0.212	0.202	-0.513	8.303	52.529*	9990*	-12.297*	363*
Stock Russia	620×10^{-6}	0.022	-0.172	0.288	0.302	14.056	7.681*	28245*	-13.290*	290*
Bond Brazil	623×10^{-6}	0.017	-0.308	0.224	-1.706	58.316	105.396***	486862*	-11.937*	815*
Bond Russia	526×10^{-6}	0.011	-0.114	0.114	-1.199	20.862	81.763*	62920*	-13.357*	696*

TABLE 6.B.3. This table shows the summary statistics for the stock and bond index returns: Stock Brazil is BOVESPA index, Stock Russia is RTSI index, Bond Brazil is JPM EMBI and Bond Russia is JPM EMBI Russia. Asymptotic p-value are shown in the brackets. *,**,*** denote statistical significance at the 1, 5 and 10 % level respectively. The Ljung-Box test statistic for serial correlation up to the 5-th order.

	Russian Stock Returns			Brazilian Stock Returns		
	Estimate	Std. Error	$\mathbb{P}(>\|t\|)$	Estimate	Std. Error	$\mathbb{P}(>\|t\|)$
μ	-2.084e-02	2.707e-04	<0.01***	-2.688e-02	2.810e-04	<0.01***
ω	1.047e-05	3.055e-06	<0.01***	1.239e-05	3.381e-06	<0.01***
α	1.555e-01	2.367e-02	<0.01***	9.862e-02	1.444e-02	<0.01***
β	8.371e-01	1.858e-02	<0.01***	8.739e-01	1.917e-02	<0.01***
υ	3.986e+00	3.093e-01	<0.01***	7.168e+00	8.337e-01	<0.01***
Log Likelihood	-8326.444			-8787.509		

	Russian Bond Returns			Brazilian Bond Returns		
	Estimate	Std. Error	$\mathbb{P}(>\|t\|)$	Estimate	Std. Error	$\mathbb{P}(>\|t\|)$
μ	-3.539e-02	5.980e-05	<0.01***	-4.828e-02	7.514e-05	<0.01***
ω	6.165e-08	2.950e-08	0.0366 *	1.866e-07	7.835e-08	0.0172 *
α	9.824e-02	1.367e-02	<0.01***	1.326e-01	1.541e-02	<0.01***
β	8.936e-01	1.013e-02	<0.01***	8.633e-01	1.148e-02	<0.01***
υ	4.557e+00	3.711e-01	<0.01***	5.142e+00	4.587e-01	<0.01***
Log Likelihood	-12113.71			-12321.91		

TABLE 6.B.4. GARCH(1,1) estimates with t-Studend distributed returns with υ degrees of freedom.

	α_1^*	α_2^*	α_1	α_2	α_0
Stock Russia vs Bond Russia	2.39 (1.92, 2.87)	1.56 (1.25,1.87)	2.69 (2.06, 3.32)	1.67 (1.28, 2.06)	1.31 (1.05, 1.56)
Stock Brazil vs Bond Brazil	2.49 (1.86, 3.11)	1.79 (1.34 2.25)	2.65 (1.98, 3.32)	1.90 (1.42, 2.37)	1.19 (0.89, 1.48)
Stock Russia vs Stock Brazil	3.01 (2.34, 3.69)	2.92 (2.26, 3.57)	3.21 (2.49, 3.93)	2.82 (2.19, 3.46)	1.48 (1.15, 1.81)
Bond Russia vs Bond Brazil	1.50 (1.21, 1.80)	1.84 (1.48, 2.20)	1.54 (1.24, 1.84)	1.88 (1.52, 2.24)	1.07 (0.87, 1.28)
Stock Brazil vs Bond Russia	2.38 (1.81, 2.95)	1.75 (1.33, 2.17)	2.56 (1.94, 3.17)	1.88 (1.43, 2.33)	1.19 (0.90, 1.47)
Stock Russia vs Bond Brazil	2.22 (1.67, 2.76)	2.01 (1.51, 2.50)	2.86 (2.15,3.56)	1.95 (1.47, 2.42)	1.01 (0.76,1.25)

	α_1^*	α_2^*	α_1	α_2	α_0
Stock Russia vs Bond Russia	2.30 (1.81 2.79)	1.98 (1.56, 2.40)	2.37 (1.82, 2.92)	2.48 (1.91, 3.06)	1.16 (0.89,1.43)
Stock Brazil vs Bond Brazil	3.57 (2.76, 4.39)	2.17 (1.67, 2.66)	3.64 (2.81, 4.47)	2.99 (2.31, 3.68)	1.11 (0.86, 1.36)
Stock Russia vs Stock Brazil	2.42 (1.83 3.01)	3.38 (2.56, 4.20)	2.54 (1.92, 3.16)	3.85 (2.91, 4.79)	1.20 (0.91, 1.50)
Bond Russia vs Bond Brazil	2.82 (2.24,3.41)	2.09 (1.66, 2.53)	2.63 (2.13, 3.13)	1.96 (1.58, 2.33)	1.43 (1.13, 1.71)
Stock Brazil vs Bond Russia	3.45 (2.56, 4.34)	2.65 (1.97, 3.33)	3.54 (2.74, 4.33)	3.04 (2.36, 3.73)	1.11 (0.86, 1.36)
Stock Russia vs Bond Brazil	2.21 (1.67, 2.76)	2.01 (1.51, 2.49)	2.93 (2.32, 3.53)	2.21 (1.75, 2.67)	1.55 (1.26, 1.85)

TABLE 6.B.5. Tail index estimates for daily stock and bond market returns over the first (top panel) and second (bottom panel) period of study, where the first two columns from rigth are the estimations for the returns not filtered for heteroskedasticidy, while the third and fourth column are the estimations for the GARCH filtered residuals. The last column is the estimation for the hidden tail index after the standarization of the marginals. Numbers presented in parentheses are standard errors.

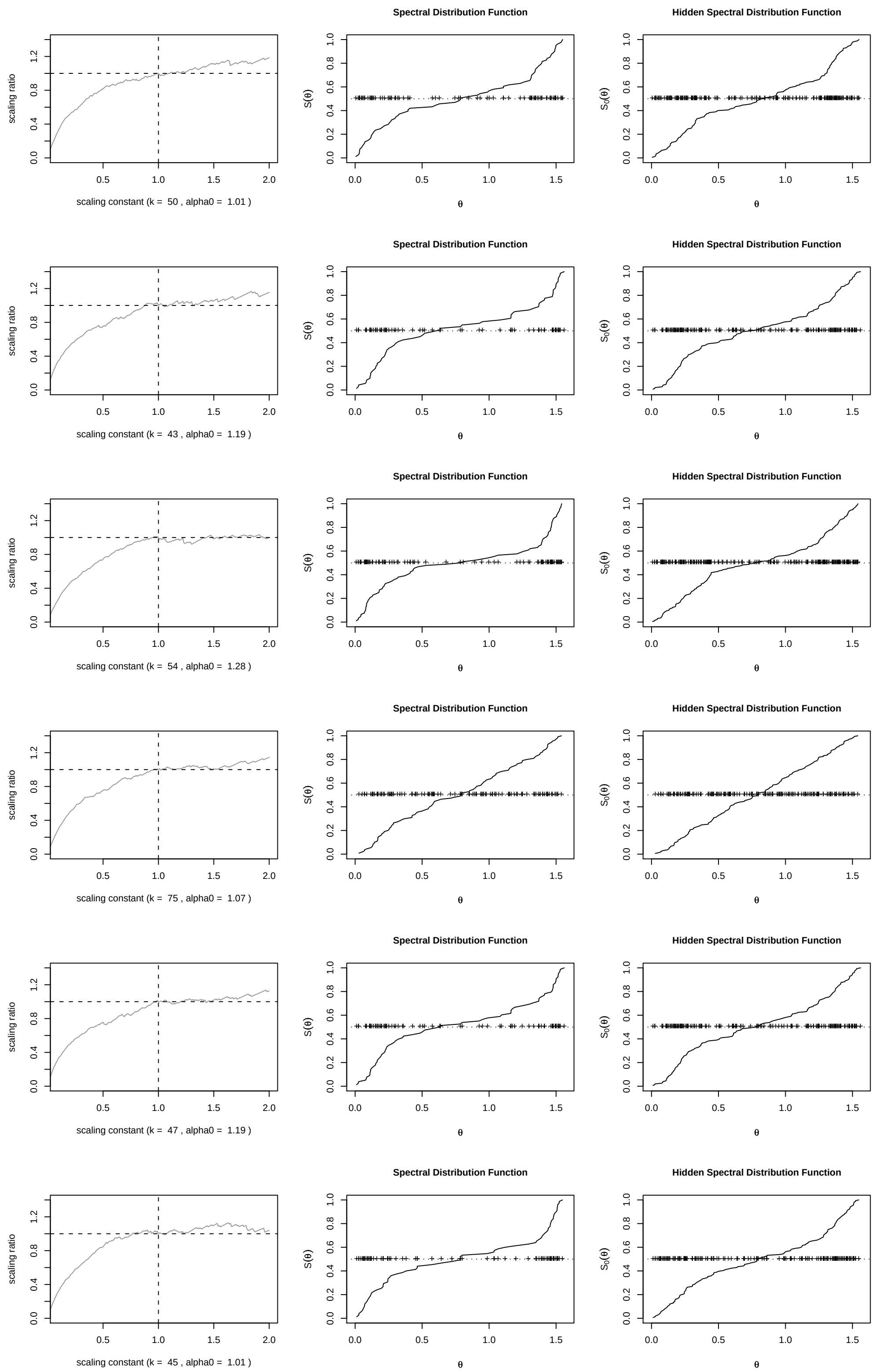

FIGURE 6.B.1. Estimations of the Starica plot (rigth), the distribution of Spectral measure (middle) and the hidden spectral measure (left) for the optimal number of exceedances (k) for the first period of study of the empirical analysis.
From top to bottom: Bond Russia vs Stock Russia, Bond Brazil vs Stock Brazil, Stock Brazil vs Stock Russia, Bond Brazil vs Bond Russia, Bond Russia vs Stock Brazil, Bond Brazil vs Stock Russia.

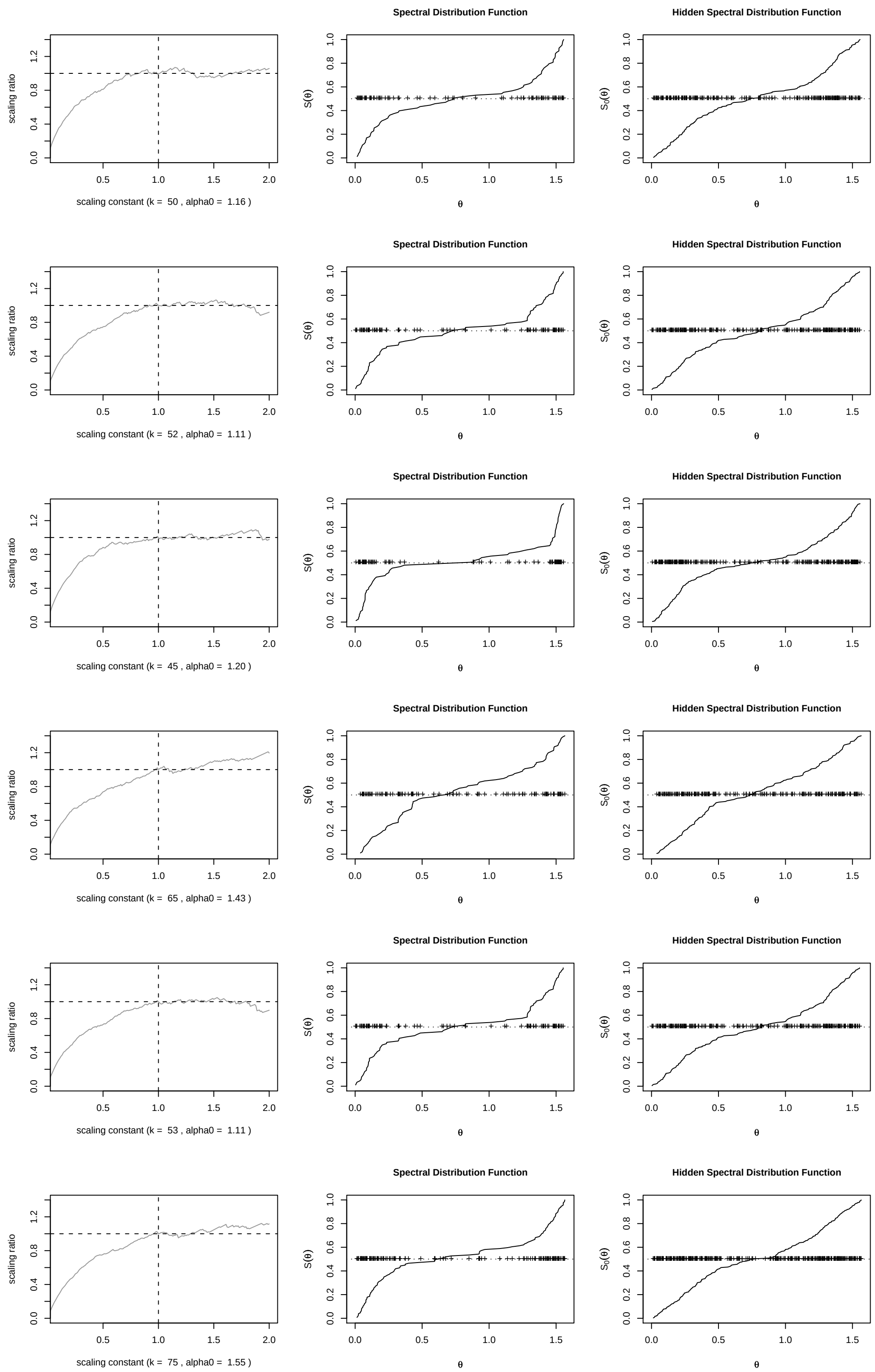

FIGURE 6.B.2. Estimations of the Starica plot (rigth), the distribution of Spectral measure (middle) and the hidden spectral measure (left) for the optimal number of exceedances (k) for the second period of study of the empirical analysis.
From top to bottom: Bond Russia vs Stock Russia, , Bond Brazil vs Stock Brazil, Stock Brazil vs Stock Russia, Bond Brazil vs Bond Russia, Bond Russia vs Stock Brazil, Bond Brazil vs Stock Russia.

Bibliography

Arestis, P., G. Caporale, and N. Spagnolo (2005). Testing for financial contagion between develop and emerging markets during the 1997 east asian crisis. *Brunel Business School Economics and Finance Working papers 5.*

Baig, T. and I. Goldfajn (1999). Financial market contagion in the asian crisis. *IMF Staff Papers 46*, 167–195.

Balkema, G. and P. Embrechts (2007). *High risk scenarios and extremes.* European Mathematical Society.

Beirlant, J., Y. Goegebeu, J. Segers, J. Teugels, and D. D. Waal (2004). *Statistics of Extremes: Theory and Applications.* Wiley Series in Probability and Statistics.

Bekaert, G. and C. Harvey (2005). Time-varying world market integration. *Journal of Business 78(1)*, 39–69.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation.* Encyclopedia of Mathematics and its Applications (No. 27).

Blackwell, D. and J. MacQueen (1973). Ferguson distributions via Polya urn schemes. *Ann. Statist 1*(2), 353 – 355.

Blei, D. and M. Jordan (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis 1*(1), 121–144.

Boldi, M. and A. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(2), 217–229.

Boyer, B. H., M. S. Gibson, and M. Loretan (1999). Pitfalls in tests for changes in correlations. *International Finance Discussion Papers 597.*

Breidt, F. J. and R. A. Davis (1998). Extremes of stochastic volatility models. *The Annals of Applied Probability 8 No.3*, 664–675.

Caillault, C. and D. Guegan (2005). Empirical estimation of tail dependence using copulas. application to asian markets. *Quantitative Finance 5*, 489–501.

Calvo, S. and C. Reinhart (1996). Capital flows to latin america: Is there evidence of contagion effects? *in Guillermo A. Calvo, Morris Goldstein, Eduard Hochreiter (eds.) Private Capital Flows to Emerging Markets After the Mexican Crisis, (Washington, DC: Institute for International Economics)*, 151–171.

Caporale, G., A. Cipollini, and N. Spagnolo (2003). Testing for contagion: a conditional correlation analysis. *Journal of Empirical Finance 12*, 476–489.

Capéraà, P., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika 3*, 567–577.

Chamú, F. (2005). *Estimation of Max-Stable Processes Using Monte carlo Methods with Applications to Financial Risk Assesment.* Ph. D. thesis, University of North Carolina at

Chapel Hill.

Chavez-Demoulin, V., A. Davison, and A. McNeil (2005). A point process approach to value-at-risk estimation. *Quantitative Finance 5 (2)*, 227–234.

Clark, P. (1973). A subordinated stochastic process model with finite variances for speculative prices. *41*, 135–156.

Coles, S.G., J. H. and J. Tawn (1999). Dependence measures for extreme value analysis. *Extremes 2 (4)*, 339–365.

Coles, S. (2001). *An introduction to Statistical Modeliing of Extreme Values*. Springer.

Coles, S. and J. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Society 53*, 377–392.

Coles, S. G. and J. A. Tawn (1994). Statistical methods for multivariate extremes: an application to structural design. *Applied Statistics, 43*, 1–48.

Corsetti, G., M. Pericoli, and M. Sbracia (2005). Some contagion, some interdependence: More pitfalls in tests of financial contagion. *Journal of International Money and Finance 24*, 1177–1199.

Csörgö, M. and L. Horváth (1997). Limit theorems in change-point analysis. *Wiley, Chichester*.

Daley, D. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes*. Springer Series in Statistics.

Das, B. and S. I. Resnick (2008). Conditioning on an extreme component: Model consistency and regular variation on cones. *Submitted*.

Davis, R. and T. Mikosch (1998). The sample acf of heavy-tailed stationary processes with applications to arch. *Annals of Statistics 26*, 2049–2080.

Davis, R. and T. Mikosch (2006a). Extreme value theory for garch processes. *(Submitted to Handbook of Financial Time Series)*.

Davis, R. and T. Mikosch (2006b). Extremes of stochastic volatility models. *(Submitted.)*.

De Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 1194–1204.

de Haan, L. and J. de Ronde (1998). Sea and wind: Multivariate extremes at work. *Extremes 1*, 7–45.

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

De Vries, C. (2005). The simple economics of bank fragility. *Journal of Banking and Finance 29*(4), 803–825.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis 13*, 257–272.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Dias, A. and P. Embrechts (2004). Dynamic copula models for multivariate high-frequency data in finance. *Warwick Business School, Finance Group, University of Warwick.*.

Dungey, M., R. Fry, B. González-Hermosillo, and V. L. Martin (2005). Empirical modelling of contagion: a review of methodologies. *Quantitative Finance 5*, 9–24.

Dungey, M., R. Fry, and V. Martin (2006). Correlation, Contagion, and Asian Evidence. *Asian Economic Papers 5*(2), 32–72.

Eichengreen, B., A. Rose, and C. Wyplosz (1998). Contagious currency crisis. *Scandinavian Economic review 4*, 463–84.

Einmahl, J., L. Haan, and V. Piterbarg (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Annals of Statistics 29*, 1401 – 1423.

Embrechts, P., A. Höing, and A. Juri (2003). Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics 7*, 145–167.

Embrechts, P., A. Höing, and G. Puccetti (2005). Worst VaR scenarios. *Insurance Mathematics and Economics 37*(1), 115–134.

Embrechts, P., C. Kluppelberg, and T. Mikosch (1997). *Modelling Extremal Events.* Springer.

Embrechts, P., F. Lindskog, and A. McNeil (2003). Modelling dependence with copulas and applications to risk management. *In Handbook of heavy tailed distributions in finance, Rachev S.T. (ed.). Amsterdam: Elsevier/North-Holland.*, 329 – 384.

Embrechts, P., A. McNeil, and D. Straummann (2002). *Correlation and Dependency in Risk Management: Properties and Pitfalls.* In: Risk Management: Value at Risk and Beyond, Cambridge: Cambridge University Press.

Engle, R. (1982a). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*, 987–1006.

Engle, R. F. (1982b). Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation. *Econometrica 50*, 987–1008.

Escobar, M. and M. West (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association 90*(430), 577 – 588.

Falk, M., J. Hüsler, and R. R. (2004). *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser, Basel.

Ferguson, T. (1983). By mixtures of normal distributions. *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*.

Ferro, C. and J. Segers (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society, Series B 65*, 545–556.

Finkenstädt, B. and H. Rootzén (2004). *Extreme values in finance, telecommunications, and the environment.* Chapman & Hall/CRC.

Forbes, K. and R. Rigobón (2001). Contagion in latin america: Definitions, measurement, and policy implications. *Economía 1*, 1–46.

Forbes, K., R. R. (2002). No contagion, only interdependence: measuring stock market comovements. *Journal of Finance 57*, 2223–2261.

Frahm, G. (2006). On the extremal dependence coefficient on multivariate distributions. *Statistics and Probability Letters 76*, 1470–1481.

Fratzscher, M. (2002). On currency crises and contagion. *European Central Bank Working Paper Series 139.*

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association 70*, 674 – 680.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Krieger Publishing Company.

Geluk, J., L. De Haan, and C. De Vries (2007). Weak & strong financial fragility.

Giacomini, E., W. Härdle, and V. Spokoiny (2007). Inhomogeneous dependency modelling with time varying copulae. *Journal of Business and Economic Statistics*.

Gombay, E. and L. Horváth (1996). On the rate of approximations for maximum likelihood tests in change-point models. *Journal of Multivariate Analysis 56*, 120–152.

Haan, L. D. and S. Resnick (1998). On asymptotic normality of the hill estimator. *Communications in statistics. Stochastic models 14*, 849–866.

Hartmann, P., S. Straetmans, and C. Vries (2004). Asset market linkages in crisis periods. *Review of Economics and Statistics 86*(1), 313–326.

Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*, 379–402.

Hefferman, J. E. and S. I. Resnick (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability 17*, 537–571.

Heffernan, J. and S. Resnick (2005). Hidden regular variation and the rank transform. *Advances in Applied Probability 37*, 393–414.

Heffernan, J. and J. Tawn (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B 66*, 497–546.

Heffernan, J., J. Tawn, and Z. Zhang (2007). Asymptotically (in) dependent multivariate maxima of moving maxima processes. *Extremes 10*(1), 57–82.

Herrera, R., M. R. Farzanegan, and A. Karmann (2008). Tail dependence in a multivariate regular variation framework: A note on financial contagion from russia to latin america. *Submitted*.

Herrera, R. and B. Schipp (2009). Self-exciting extreme value models on stock market crashes. *In : Statistical Inference, Econometric Analysis and Matrix Algebra. Editors : Bernhard Schipp and Walter Krämer. Physica, Heidelberg*, 209–231.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics 21*, 2043–2071.

Hsing, T., J. Hüsler, and M. Leadbetter (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields 78*, 97–112.

Hsing, T., C. Klüppelberg, and G. Kuhn (2004). Dependence estimation and visualization in multivariate extremes with applications to financial data. *Extremes 7*, 99–121.

Hüsler, J. and D. Li (2008). Testing asymptotic independence in bivariate extremes. *Journal of Statistical Planning and Inference*.

Huang, X. (1992). *Statistics of bivariate extremes*. Ph. D. thesis, Erasmus Univ. Tinbergen Institute Research Series.

Jalal, A. and M. Rockinger (2004). Predicting tail-related risk measures: The consequences of using garch filters for non-garch data," fame research paper series. *FAME Research Paper Series, International Center for Financial Asset Management and Engineering 115.*

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. & Probab. Letters 9*, 75–81.

Joe, H. (1993). Parametric family of of multivariate distributions with given margins. *Journal of the Multivariate Analysis 46*, 262–282.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Monographs on Statistics and Applied Probability No. 37, Chapman & Hall.

Jorion, P. (2003). *Financial risk manager handbook.* Wiley.

Kagan, Y. Y. and L. Knopoff (1987). Statistical short-term earthquake prediction. *Science 235*, 1563–1467.

Klüppelberg, C. and A. Mayer (2006). Bivariate extreme value distributions based on polynomial dependence functions. *Mathematical methods in the Applied Sciences 29*, 1467 –1480.

Kluppelberg, C. (1998). Subexponential distributions and integrated tails. *Journal of Applied Probability 25 (1)*, 132–141.

Kole, E. and E. U. Rotterdam (2006). *On crises, crashes and comovements.* Erasmus Universiteit Rotterdam.

Kotz, S. and S. Nadarajah (2002). *Extreme Value Distributions: Theory and Applications.* Imperial College Press.

Krugman, P. (1999). Don't blame it on rio ... or brasilia either. *Slate magazine*.

Kurihara, K., M. Welling, and N. Vlassis (2007). Accelerated variational dirichlet process mixtures. *Advances in Neural Information Processing Systems 19*, 761.

Laurini, F. and J. Tawn (2003). New estimators for the extremal index and other cluster characteristics. *Extremes 6*, 189–211.

Leadbetter, M., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer.

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

MacEachern, S. and P. Mueller (1998). Estimating Mixture of Dirichlet Process Models. *Journal of Computational and graphical Statistics 7*, 223 – 238.

Malevergne, Y. and D. Sornette (2005). *Extreme Financial Risks. From Dependence to Risk Management.* Springer, Berlin.

Markowitz, H. (1991). *Portfolio selection: efficient diversification of investments.* Blackwell Publishing.

Martins, A. and H. Ferreira (2005). Measuring the extremal dependence. *Statistics and Probability Letters 73*, 99–103.

Matthys, G. and J. Beirlant (2001). Extreme quantile estimation for heavy tailed distributions. *Universitair centrum voor Statistiek, Katholieke Universiteit Leuven.*.

Maulik, K. and S. Resnick (2004). Characterizations and examples of hidden regular variation. *Extremes 7*, 31–67.

McKibbin, W. and W. Martin (1998). The east asian crisis: Investigating causes and policy responses. *Working Papers in Trade and Development, Australian National University*.

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

McNeil, A. J. (1998). On extremes and crashes. *Risk 11*, 99–104.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, Princeton.

Meerschaert, M. and H. Scheffler (2001). Sample cross-correlations for moving averages with regularly varying tails. *Journal of Time Series Analysis 22*, 481–492.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes 9*(1), 3–20.

Moser, T. (2003). What is international financial contagion? *International Finance 6*, 157–178.

Nadarajah, S. (2001). Multivariate declustering techniques. *Envirometrics 12*, 357–365.

Nandagopalan, S. (1994). On the multivariate extremal index. *Journal of Research of the National Institut of Standard and Technology 99*, 543–550.

Neal, R. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and graphical Statistics 9*(2), 249 – 265.

Nelsen, R. (2006). *An Introduction to Copulas*. Springer Series in Statistics.

Novak, S. (2004). Value at risk and the "black monday" crash. *Accounting & Finance*.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association 83*, 9–27.

Omey, E. (1994). On the difference between the product and the convolution product of distribution functions. *Lithuanian Mathematical Journal 38*, 1–14.

Omey, E., F. Mallor, and J. Santos (2006). Multivariate subexponential distributions and random sums of random vectors. *Advances in Applied Probability 38 (4)*, 1028–1046.

Patton, A. J. (2006). Modelling ssymmetric exchange rate dependence. *International Economic Review 47*, 527–556.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006, March). CODA: Convergence diagnosis and output analysis for MCMC. *R News 6*(1), 7–11.

Poon, S., M. Rockinger, and J. Tawn (2001). New extreme-value dependence measures and finance applications. *CEPR Discussion Paper 2762*.

Poon, S.-H., M. Rockinger, and J. Tawn (2003). Modelling extreme-value dependence in international stock markets. *Statistica Sinica 14*, 929 – 954.

R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rasmussen, C. (2000). The infinite Gaussian mixture model. *Advances in Neural Information Processing Systems 12*, 554 – 560.

Reiss, R. (1989). *Approximate Distributions of Order Statistics: With Applications to Non-parametric Statistics*. Springer Series in Statistics.

Resnick, S. (1987). Extreme values, regular variation, and point processes. *Springer, New York*.

Resnick, S. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, Berlin.

Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance 14*, 401–423.

Rosenblatt, M. (1971). *Markov Processes, Structure and Asymptotic behaviour*. Springer.

Schlather, M. and J. Tawn (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes 5*, 87–102.

Seger, J. (2002). Extreme events: Dealing with dependence. *EURANDOM Report 36*.

Segers, J. (2003). Functionals of clusters of extremes. *Advances in Applied Probability*, 1028 –1045.

Smith, R. (1989). Extreme value analysis of environmetal time series: An application to trend detection in ground-level ozone. *Statistics Sience 4*, 367–393.

Smith, R. and I. Weissman (1996). Characterization and estimation of the multivariate extremal index. *Unpublished*.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *pre-print. University of North Carolina, USA.*.

Smith, R. L. (2003). Statistics of extremes, with applications in environment, insurance and finance. *In : Finkenstädt and Rootzén*, 1–78.

Smith, R. L., J. A. Tawn, and S. G. Coles (1997). Markov chain models for threshold exceedances. *Biomatrika 84*(2), 249–268.

Starica, C. (1999). Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance 6*, 515–553.

Straetmans, S. (2000). Extremal spill-overs in equity markets. *Extremes and Integrated Risk Management, Risk Books (London)*, 187–204.

Straetmans, S., W. Verschoor, and C. Wolff (2008). Extreme us stock market fluctuations in the wake of 9/11. *Journal of Applied Econometrics 23*(1), 17– 42.

Sull, A. (2006). Global financial contagion identification : from russia to brazil. Master's thesis, Simon Fraser University.

Tawn, J. A. (1988a). Bivariate extreme value theory: Models and estimation. *Biometrika 3*, 397–415.

Tawn, J. A. (1988b). An extreme value theory model for dependent observations. *Journal of Hydrology 101*, 227–250.

Teh, Y., M. Jordan, M. Beal, and D. Blei (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association 101*(476), 1566–1581.

van den Goorbergh, R. W., C. Genest, and B. J. Werkerc (2005). Bivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics 37*, 101–114.

Yang, T. and J. Lim (2004). Crisis, contagion, and East Asian stock markets. *Review of Pacific Basin Financial Markets and Policies 7*(1), 119 – 151.

Yun, S. (2000). The distributions of cluster functionals of extreme events in a dth-order Markov chain. *Journal of Applied Probability*, 29 – 44.

Zhang, Z. (2002). *Multivariate Extremes, Max Stable Process Estimation and Dynamic Financial Modelling*. Ph. D. thesis, University of North Carolina at Chapel Hill.

Zhang, Z. (2008). On approximating max-stable processes and constructing extremal copula functions. *Statistical Inference for Stochastic Processes*, 1–26.

Zhang, Z. and R. Smith (2004). The behavior of multivariate maxima of moving maxima processes. *J. Appl. Prob 41*, 1113–1123.