EXTREME VALUE THEORY IN RISK MANAGEMENT

Doctoral Thesis

Author: José Olmo

Advisor: Jesús Gonzalo

Department of Economics

UNIVERSIDAD CARLOS III DE MADRID

Getafe, March 2005

This thesis is dedicated to my father, José.

Contents

"La fantasía, abandonada de la razón,

solo produce monstruos imposibles.

Unida a ella, en cambio, es la madre

del arte y fuente de sus deseos."

Francisco de Goya

"Sólo la fantasía permanece siempre joven;

lo que no ha ocurrido jamás no envejece nunca.

J. C. Friedrich von Schiller

ACKNOWLEDGEMENTS

I owe thanks primarily to my advisor, professor Jesús Gonzalo. He has taught me to be rigorous and thorough, to think and to conduct research. His suggestions and commnents have been essential in this thesis.

I gratefully acknowledge his name, and to Universidad Carlos III de Madrid its surname. This thesis would have not been possible without the support of the Departments of Economics, and specially, of the Department of Statistics and Econometrics. Their resources and faculty have been indispensable for the development of this research project. I also owe thanks to the Ph.D. program by keeping very high quality standards that have allowed me to meet international prestige professors as Rob Engle, Clive Granger, David Nualart, Winfried Stute, or Murad Taqqu. The comments of some of them have been really helpful.

I owe thanks to Fundación Universidad Carlos III scholarships for short stays abroad that gave me the opportunity of visiting Edimburgh University (summer 2000), and Oxford University (summer 2001).

This research projet has been supported by the Spanish Secretary of Science and Technology and by the Spanish Secretary of Education and Science, with the project number SEC2001-0890, from which I have participated with a training research grant during the last three years. I am grateful to this grant that has permitted me visiting the Department of Statistics and Operations Research of UNC-Chapel Hill during the last two fall semesters.

I would not like forgetting this department. Its hospitality and research environment have been fundamental for the output of this thesis. I would highlight the advise of Professors Richard Smith and Ross Leadbetter. From the Economics Department of UNC I would like to mention Professor Eric Renault.

From all these places what I have appreciated most is friendship. From Universidad Carlos III I will start by mentioning my good friends Ricardo Laborda and Pablo Villaplana. Besides, Sara López, Oscar Martínez, Ramiro Losada, Juan Carlos Escanciano, Andrés Ubierna, Mónica Benito, Carmen Broto, María Angeles Pulido and others that haven been around eventually. From North Carolina, Vassilis Maroulas, Arka Ghosh, Brian Lopes, Gustavo Didier and in particular Francisco Chamú and his wife Jennifer. From Oxford, Asghar Zaidi and Rosa Fernández, and from Edimburgh, Professor Andy Snell.

The last lines are for my family. I owe many thanks to my brother Jorge, and to my mother María Asunción because without her there would not be lagrange multipliers, martingales or second moments. To Ana, because she is the sweetest person I have ever met, and finally to my father. Though he could not see this, he always knew I would get it. He was very proud of me, but I am much more of him.

SUMMARY

The intention of this dissertation is to provide some insight about risk management by using a methodology far from the standard statistical techniques: variance and correlation. The alternative is Extreme Value Theory, that is presented as the natural setup to quantify risk in financial econometrics.

The thesis concentrates on risk. There are different interpretations of this concept that result in diverse methodologies to quantify its magnitude and impact on different characteristics of financial econometrics. In the introduction of the thesis the distinction between uncertainty and risk is discussed, regarding the point of view: decision theory or risk management. It follows with a formal definition of risk motivated by decision theory but consistent with the methodology used in risk management. Risk can be quantified by means of statistical techniques. Risk is characterized by the tails of the distribution of the data, in particular by the likelihood of any event entailing a negative feature. In financial econometrics this definition of risk is usually denominated downside risk and is associated with the left tail of the distribution of returns.

The aim of the second chapter is to provide reliable measures to quantify the risk found in financial sequences. In order to achieve this, standard tools of extreme value theory are applied.

All the risk measures recently considered in the literature based on extreme values are characterized in practice by ad-hoc selection methods for the extreme values (5%, 1%, etc.) The main contribution in the second chapter is to propose a formal definition for these values. The extreme values of any random sample of size n from a distribution function F are defined as the observations exceeding a threshold and following a type of generalized Pareto distribution (GPD) involving the tail index of F. The threshold is the order statistic that minimizes a Kolmogorov-Smirnov statistic between the empirical distribution of the corresponding largest observations and the corresponding GPD. To formalize the definition we use a semiparametric bootstrap to test the corresponding GPD approximation. Finally,

we use our methodology to quantify risk by estimating the tail index (ratio of decay of the negative tail), and the value at risk (VaR) of some financial indexes of major stock markets.

Once risk is defined and formally quantified the following aim in the thesis is analyzing its transmission mechanisms in different settings. Chapter 3 is devoted to the transmission of risk in time series. The risk is measured by the occurrence of significant large observations and the transmission channel is the serial dependence found in the extreme values that can originate clustering of data. In this context there exists a parameter, the extremal index, that governs the serial dependence in the largest observations, and such that its reciprocal measures the level of clustering in the extremes. The contribution of the thesis in this chapter starts by redefining this parameter. This definition provides a straightforward estimation method for the extremal index with appealing statistical properties; consistency, and asymptotic gaussian distribution. The existence of clustering in the largest observations is a byproduct of the transmission of risk derived from the occurrence of the largest observations. An outstanding contribution in this part is the possibility of testing the transmission of risk in financial sequences by testing the clustering in the extreme values. This theory contrasts with theories founded on volatility models that claim that serial dependence found in financial series is due to the conditional dependence on second moments.

The next chapter involves the transmission of risk between financial markets. The interest lies in this section on distinguishing interdependence between markets, that surges from regular links between economies, from contagion effects, originated by increasing links between the markets in crises periods. In order to do this, the notions of interdependence and contagion are revisited. The contribution of the authors lies on new definitions for these concepts based on copula properties and tail monotonicity, that will be used to analyze directional contagion (causality between extremes). This is possible due to an innovative copula function that is derived from the multivariate extreme value theory. This copula allows us to model different patterns of dependence according to the state of the markets, *e.g.* bear or bull markets. This model is sufficiently flexible to describe asymmetries between variables in such a way that directional contagion can be tested. The model is applied to the flight to quality phenomenon, outflows of capital from the stocks markets to the bonds markets when the first ones are facing crisis periods.

Finally Chapter 5 sketches future lines of research involving different aspects of the analysis of risk.

Chapter 1

Introduction

Summary

This chapter presents the problems and concerns that motivated this research project. The section begins with the notions of uncertainty and risk derived from decision theory and develops the corresponding statistical treatment. In the current literature, variance is used for quantifying both of them. In this part however, the differences between uncertainty and risk are discussed and the appropriate statistical methodologies for each problem are sketched. The Extreme Value Theory is motivated as the natural environment to quantify risk. The pitfalls of this technique and the alternatives presented in the thesis are introduced. The section concludes with the multivariate setting and the use of copula functions to model dependence motivated by the pitfalls of linear measures as correlation.

In the last decade Risk Management has become a major discipline in Finance. It is studied in different fields within finance: financial econometrics, mathematical finance or financial engineering. The main concern of risk management is analyzing the causes and consequences of negative events for investor interests. This rough definition depends very much on the definition of negative event, and on the profile of the investor. In addition, one of the more promising research areas in finance in the last years is the development of financial instruments and investment strategies that allow one hedging from negative events. Simple examples of the latter are the use of stock options, the possibility of positioning long or short in a portfolio of assets or investing in different derivatives with corresponding different maturities.

The concept of risk, in turn, is not clearly defined. At least, there is not a universal definition that permits academics and practitioners to progress in the same direction to solving the same problems. Both groups however, have concentrated on studying the uncertainty rather than risk guided by the common belief that both terms are exchangeable. Knight (1921) defined uncertainty of occurrence of a particular event by the impossibility of assigning a probability to the event. In this way this author defined uncertainty situations by the absence of insurance markets. In contrast, risk is present when we can assign a probability measure to the event, and therefore insurance markets develop (trade exists).

One other important distinction between these concepts is given by the negative feature of the event producing risk. Uncertainty does not necessarily entails a negative outcome. On the contrary, risk implies a strictly positive likelihood of a negative result in the universe of possible outcomes. Both concepts however, are characterized by the presence of randomness and therefore can be attributed to a random variable. In this way, the former definitions for risk and uncertainty can be translated to probability theory.

In this context, uncertainty is intrinsic to the definition of random variable and is usually described by the variance. Risk however entails something more not captured by the variance. Consider for example the forecast of the weight of an adult person when he is ten years old. No doubt that statistical uncertainty is measured by the variance, albeit the latter is not informative at all about the risk. Risk in this situation comes from very low or high forecasted values. In this example the extreme observations can derive in different illnesses: anorexia, obesity.

This example raise the issue of defining risk as something occurring in the tails of the distribution of the random variable and entailing the knowledge of its probability distribution. It is interesting however the statisticians and econometricians vision of risk. It boils down to measuring the variance of the random variable describing the event. This is only true if the probability distribution is known and the only unknown is the variance. Consider the example of a normal distribution $N(\mu, \sigma^2)$ where μ is unknown but σ is known. The knowledge of σ is not sufficient to know the uncertainty neither the risk. The probability distribution is not known, but a set of possible distribution functions. This unusual example reflects the ambiguity of knowing only the variance. In the particular example of analyzing financial returns it is a common hypothesis to assume the expected value of the returns to be zero, and then makes sense to think of the volatility (variance) as measuring the risk. Nevertheless this example derived from financial econometrics needs of another assumption. The prices of the financial instrument (asset, bond) are assumed to follow a log-normal distribution, and in consequence the returns distribution is assumed normal. Other situation where the variance is sufficient to describe the overall risk of financial returns is when the preferences function of the investor (utility function) is quadratic.

There is a handful of econometric techniques for estimating the risk under these assumptions. The focus is in the estimation and modelling of the volatility process. The standard methodology is estimation from the historical distribution where the volatility is considered constant, and all the observations have the same weight in estimating the variance. Instead, if some dynamics is observed in the data, a more adequate estimator for the volatility is some exponential smoothing technique where the most recent observations $\{x_t\}$ have more protagonism than past observations.

$$\sigma_t^2 = (1 - \lambda)x_{t-1}^2 + \lambda\sigma_{t-1}^2, \tag{1.1}$$

with $0 < \lambda < 1$.

The same philosophy is followed by GARCH models introduced in Engle and Bollerslev (1986). The general GARCH(p,q) model takes this form

$$\sigma_t^2 = \omega + \alpha_1 x_{t-1}^2 + \ldots + \alpha_p x_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2, \qquad t = 1, \ldots, T \tag{1.2}$$

with $\omega, \alpha_i, \beta_j > 0$ and $\sum_i \alpha_i + \sum_j \beta_j < 1$ to obtain stationarity.

There are minor modifications of this model reflecting different stylized facts of the financial data. Examples of these models are EGARCH regarding the leverage effect, IGARCH where $\sum_i \alpha_i + \sum_j \beta_j = 1$ describing infinite variance (infinite risk?), etc.

More sophisticated forms of measuring risk in this setting are given by the implied volatility and the realized volatility, that have been developed in the last years. Implied volatility is derived from option pricing and in consequence from Black-Scholes formula, Black and Scholes (1973). The prices are supposed to follow a geometric Brownian Motion. This assumption is sufficient, for example, to find the no-arbitrage price for the European plain-vanilla option,

$$C_t(S_t) = S_t \Phi(d_1) - K e^{-r(T-t)} K \Phi(d_2), \tag{1.3}$$

with C_t the option price, $\Phi(\cdot)$ the standard normal distribution function, r the risk free interest rate, K the strike, T the maturity, and d_1, d_2 constants satisfying

$$d_1 = \frac{log \frac{S_t}{K} + (r + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}}, \qquad d_2 = d_1 - \sigma \sqrt{T-t}.$$

The knowledge of the option prices for different maturities is observable from the market. Therefore, from equation (1.3) the variance can be obtained. This variance, denoted implied volatility, is derived in turn, from agents expectations about the future.

Other volatility measure founded on stochastic differential equations is the realized volatility. This concept is originated in papers by Andersen et al (2001), and Barndorff-Nielsen and Shephard (2002). The expression for volatility builds on the theory of continuous-time arbitrage-free price processes and the theory of quadratic variation. Denote $[x]$ for the quadratic variation of x defined as

$$[x]_t = \underset{q\to\infty}{plim} \sum_{i=0}^{m-1} \{x_{t_{i+1}^q} - x_{t_i^q}\}^2, \tag{1.4}$$

for any sequence of partitions $t_0^q = 0 < t_1^q < \ldots < t_m^q = t$ with $sup_i(t_{i+1}^q - t_i^q) \to 0$ for $q \to \infty$. The notation *plim* indicates the probability limit of the sum. If the log of prices follow a stochastic differential equation of this type

$$x_t = \mu_t dt + \sigma_t dw_t, \tag{1.5}$$

with σ_t^2 the instantaneous or spot volatility, the quadratic variation takes the same expression than the integrated volatility defined as

$$\sigma_t^{2*} = \int_0^t \sigma_u^2 du. \tag{1.6}$$

This method explodes the availability of high frequency intra-period returns. The realized volatility, denoted $\{x\}_n$ is the natural estimator of the quadratic variation, and is defined as the sum of the squared returns of M intra-day observations during each day. It takes this expression

$$\{x\}_n = \sum_{j=1}^{M} \left\{ x_{(n-1)+\frac{j}{M}} - x_{(n-1)+\frac{j-1}{M}} \right\}^2, \tag{1.7}$$

that is consistent as $M \to \infty$. The theory of quadratic variation reveals that under suitable conditions, realized volatility is not only an unbiased ex-post estimator of daily return volatility, where the day index is n, but also asymptotically free of measurement error. Empirically, by treating the volatility as observed rather than latent facilitates modelling by using simple methods based directly on observable variables. On the other hand volatility regarded as a latent variable, not observed, may be modelled by stochastic volatility models. These models are of this form

$$\sigma_t = \rho\, \sigma_{t-1} + \epsilon_t, \tag{1.8}$$

with $0 < \rho < 1$ the autoregressive parameter, and ϵ_t the innovation variable.

All of the above different methodologies to quantify risk fail if the distribution of returns is far from the gaussian assumption. This fact is gaining popularity within the academics and practitioners that have raised the need of a more realistic modelling of the distribution of returns, and of the analysis of risk. The focus moves from a measure for the dispersion of the data to a measure that describes the probability in the tails. The risk underlying the financial sequence is renamed as downside risk since it is associated to negative outcomes that are usually represented in the left tail of the distribution of returns. It is worth mentioning the upside risk due to the existence of hedging instruments that are designed to compensate values in the left tail and can yield negative outcomes when the returns take on large positive values. The interest of risk managers is found in estimating the distribution of the data, in particular the distribution in the tails. The results found in Kolmogorov (1933) and in Gnedenko (1943) derived from the distribution of the sample maximum are the basis of a new and exciting area in Statistics involving the analysis of the extreme values of random sequences and the distribution in the tails. This area is denominated Extreme Value Theory (EVT) and is the theoretical basis and statistical toolkit for the techniques developed in this thesis.

Suppose a random sample from an unknown distribution function F, and let G be the limiting distribution of the sample maximum M_n. Classical Extreme Value Theory shows that under some regularity conditions on the tail of F and for some suitable constants $a_n > 0$, b_n,

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x), \tag{1.9}$$

where G must be of the following types (see de Haan (1976)),

Type I: (Gumbel) $\quad G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty.$

Type II: (Fréchet) $\quad G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\frac{1}{\xi}}} & x > 0, \xi > 0. \end{cases}$

Type III: (Weibull) $\quad G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(-x)^{-\frac{1}{\xi}}} & x < 0, \xi < 0. \end{cases}$

The parameter ξ is the tail index of F and characterizes the tail behavior of the distribution function. The three types can be gathered in the so called Generalised Extreme Value Distribution, first proposed by von Mises (1936),

$$G(x) = e^{-(1+\xi\frac{x-\nu}{\sigma})^{-\frac{1}{\xi}}}, \tag{1.10}$$

where ν is a location parameter, σ a scale parameter and $\xi \neq 0$. This expression boils down to $G(x) = e^{-e^{-\frac{x-\nu}{\sigma}}}$ when $\xi = 0$.

In consequence the distribution of the standardized sample maximum $F^n(a_n x + b_n)$ converges to $e^{-(1+\xi\frac{x-\nu}{\sigma})^{-\frac{1}{\xi}}}$ for all x. Some simple algebra yields

$$\lim_{n\to\infty} n\left(1 - F(a_n x + b_n)\right) = \left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}. \tag{1.11}$$

This result is exploited to derive the weak convergence of the largest observations determined by a threshold sequence $u_n = a_n\nu + b_n$, with ν satisfying $-log\ G(\nu) = 1$. This is the main result of Pickands theorem, Pickands (1975),

$$\underset{u_n \to x_F}{lim} \sup_{[0\leq y<\infty]} |F_{u_n}(y) - GPD_{\xi,\sigma(u_n)}(y)| = 0, \tag{1.12}$$

with

$$F_{u_n}(y) = \frac{F(u_n + y) - F(u_n)}{1 - F(u_n)}, \tag{1.13}$$

$y = a_n(x-\nu)$, $\sigma(u_n) = \sigma a_n$ and $x > \nu$. $F_{u_n}(y)$ is the conditional excess distribution function given u_n, and

$$GPD_{\xi,\sigma(u_n)}(y) = \begin{cases} 1 - (1+\xi\frac{y}{\sigma(u_n)})^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma(u_n)}} & \text{if } \xi = 0 \end{cases}, \tag{1.14}$$

is the Generalized Pareto distribution.

Pickands theorem holds promise for accurate estimation of extreme quantiles and tail probabilities of financial returns when the distribution F is unknown. In this way EVT irrupted in financial econometrics as an omnibus technique that overcame the problems derived from the absence of information in the tails. This methodology had been implemented before in other sciences as hydrology or meteorology with relative success. After some euphoria expressed by a number of papers using EVT techniques in the middle nineties some disappointment surged in academics due to pitfalls in the development of the statistical theory, and in practitioners due to its complexity in comparison with historical simulation or with the methods derived from gaussian assumptions on the data. In this line there is a paper of Diebold, Schuerman and Stroughair (1998) that enumerates some pitfalls and challenges in EVT that lie ahead. The final recommendation of these authors is some caution in the use of this methodology in financial econometrics and risk management, and a better understanding of the situations where these powerful probabilistic and statistical techniques are reliable and really helpful.

Much of the discussion in Diebold et al. is related to the statistical aspects of the theory, in particular tail estimation and the GPD approximation. These authors assume that the tail of F has a power law (decays polynomially) and in turn belongs to the maximum domain of attraction of a Fréchet distribution, that is, the distribution of the sample maximum of F converges weakly to that type of EVT distribution. The tail index estimator considered is Hill estimator, Hill (1975),

$$\hat{\xi}_n^{Hi}(u_n) = \frac{1}{k} \sum_{i=n-k+1}^{n} \log \frac{x_{(i)}}{x_{(n-k)}}, \tag{1.15}$$

with $u_n = x_{(n-k)}$, the threshold sequence and $x_{(n-k+1)} \leq \ldots \leq x_{(n)}$ denoting the increasing order statistics.

This estimator has appealing theoretical properties. It is consistent and asymptotically normal, assuming the data are independent and identically distributed (iid) and that k, the number of largest observations defined by the threshold $u_n = x_{(n-k)}$, grows at a suitable rate, $k \to \infty$ with $k/n \to 0$. It can be shown that the selection of the threshold for the Hill estimator affects its bias and variance. In particular, there is an important bias-variance tradeoff when varying k for fixed sample data. Diebold et al. point to the threshold selection problem as the first pitfall in EVT regarding estimation in the tails. They claim for a formal and rigorous procedure to determine the threshold rather than ad-hoc rules of thumb based on graphical methods. The formal rule to decide the threshold could be supplemented and assessed by bootstrap resampling techniques and Monte-Carlo simulation.

The second pitfall regarding EVT is the absence of serious studies about the finite sample properties of the tail index estimator under various threshold choices. They maintain that an intensive Monte-Carlo study is needed for various generating processes. Other related pitfall not mentioned in Diebold et al. refers to the use of the asymptotic normal approximation for the distribution of the tail index estimator for small sample sizes. The authors remark the overall poor treatment of the statistical aspects of EVT in contrast to the probabilistic results. Some important examples are the absence of confidence intervals for quantile and tail estimates and the absence of reliable hypothesis tests for small sample sizes.

The second chapter of this thesis gathers these concerns and develops a methodology overcoming most of the pitfalls attributed to EVT. The chapter presents a definition for the extreme values of a random sequence. This definition is given by the observations exceeding a threshold sequence determined by the asymptotic property in Pickands theorem. In consequence the choice of the threshold turns crucial. A formal iterative procedure for threshold selection is implemented on the basis of the GPD approximation. The threshold choice depends on the data (it is an order statistic) and on the length of the sequence. Therefore to formalize the definition of extreme values we introduce an innovative bootstrap hypothesis test. The bootstrap is semi-parametric benefiting of the historical information for the bulk of the data and from the parametric GPD fit in the tail. The obtained bootstrap sampling distribution makes immediate the study of the sample properties of the different tail index estimators considered and the statistical inference (confidence intervals for quantile and tail measures). The main beneficiaries of the accurate estimation methods obtained are risk measures. In particular VaR and tail thickness that are estimated for major financial indexes worldwide, and assessed by bootstrap confidence intervals.

The maintained assumptions in Diebold et al. are the polynomial law in the decay of the tails and the *iid* data. In consequence, departures from these assumptions will lead to more pitfalls in the statistical aspects of EVT. In particular the *iid* assumption for financial econometrics data is not realistic. The availability of high frequency data improves and motivates the use of EVT, but on the other hand entails serial dependence. The typical financial literature explains this dependence by the correlation in the second moments. In other words, while the returns are uncorrelated, the conditional volatility is driven by past information. Along with the dependence in volatility there are other stylized facts characterizing financial returns. These are the clustering of the largest observations in both tails, the magnitude of the largest observations that is far from being explained by gaussian models (heavy tails ?), the asymmetry observed in the distribution of returns, and the leverage effect, that is, after periods of high volatility periods of negative returns. A first glance to this type of sequences would promptly discard the use of gaussian assumptions, however practitioners and researchers have kept on proposing models to describe the dynamics of these data by modelling volatility.

The third chapter puts together the statistical challenge of EVT regarding the dependence in the extremes and the financial econometrics problem regarding the modelling of these stylized facts. The contributions of the chapter are twofold. First, it proposes a new estimator for the parameter describing the dependence in the extremes, denoted extremal index (θ), and second regards the clustering in the financial sequences as a characteristic of the largest observations rather than from the second moments of the conditional distribution. The statistical properties of the estimator for θ are very appealing. In particular, it is asymptotically normal making the inference straightforward. In turn, the clustering in the largest observations may be tested. Other hypothesis tests derived from the asymptotic distribution of the estimator are developed with immediate application to testing the remaining stylized facts found in financial sequences. The conclusions agree with the theories considering the presence of heavy tails in financial sequences rather than being conditionally gaussian with heteroskedastic volatility.

A major challenge in EVT is the multivariate setting. There is no natural extension of this theory to several variables. Moreover, the typical problems encountered for higher dimensions, serve as example the curse of dimensionality, are augmented in the context of extreme values. Therefore, though the applications in the real world are endless, in particular in financial econometrics and risk management for modelling dependence between markets in crises periods, these are hindered by the lack of data and rigorous techniques for dimensions higher than two.

The fourth chapter concentrates on these links stressing the contagion phenomenon, that is, the transmission of crises from ill economies to healthy economies. This has been measured so far by linear measures as correlation. This methodology is not designed to measure causality between the variables and in consequence fails in describing the source and strength of the dependence in distress periods. Embrechts, McMeil and Straumann (1999) review the pitfalls of the Pearson correlation and present some alternatives to this standard measure of dependence. The message is that Pearson correlation is an adequate measure for dependence only for the multivariate normal distribution. In this case the marginal distributions and the correlation matrix are sufficient to describe the overall dependence in the data.

On the basis of correlation, two multivariate distribution functions with the same correlation matrix cannot be differentiated, and more important for our case, correlation tells us nothing about the degree of dependence in the tails of the underlying distributions. Other problems of correlation are enumerated in the following list.

1. Correlation is simply an scalar measure of dependency. It is not designed to describe the complete structure of dependence.

2. Correlation depends on the marginal distributions. All values between -1 and 1 are not necessarily attainable.

3. Perfect positively dependent random variables do not necessarily have a correlation of 1. Perfect negatively dependent random variables do not necessarily have a correlation of -1.

4. A correlation of zero does not indicate independence between the variables.

5. Correlation is not invariant under transformations of the risks.

6. Correlation is only defined when the variances of the corresponding variables are finite.

Embrechts et al. recommend the copula functions, a novel method in economics, to model dependence. This method lies on the same idea raised in the discussion of the notions of variance and risk. While variance was a linear measure describing the dispersion of data, risk needed of the whole distribution function. In the multivariate setting, the role of variance is assumed by correlation, and the notion of risk is replaced by contagion and the strength of the dependence in the tails. Therefore, in the same way that risk needed of F, contagion needs of the whole multivariate function. The copula function extracts the dependence structure from the joint distribution function. Sklar (1959) showed that every joint distribution function can be written as

$$H(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)), \tag{1.16}$$

with H the multivariate distribution function, F_i, $i = 1, \ldots, m$ the margins, C the copula function, and m the number of random variables.

As everything in life, this powerful statistical tool is not free from pitfalls. There is an enormous set of available copulas, but there is no formal method to discriminate between them. Goodness of fit tests in the multivariate case are not straightforward, and depend heavily on the knowledge of the marginal distributions. In addition most of the techniques for testing copula fitness face computational burdens that make the practical implementation difficult. The choice of the copula is usually replaced by ad-hoc methods. The other important deficiency of the majority of the copulas is symmetry. Copula functions are usually symmetric and therefore are not designed to reflect different contributions of the corresponding variables to the dependence. Finally, the dynamics in cross dependence for multivariate data are not very much explored. Conditional copulas introduced in Patton (2001) are a first attempt to model time-varying dependencies, albeit these copulas boil down to impose some dynamics to the parameters driving the dependence in standard copulas. On the other hand the conditioning assumptions are not well specified in general.

In Chapter 4 the multivariate dependence between different financial markets is divided in two groups: rational and irrational dependence. The rational dependence is due to economic fundamentals, and is described by univariate regression models and volatility filters. The irrational dependence is reflected in the links in the vector of innovations and is modelled by copula functions. The emphasis is placed on the links between financial bonds markets and stocks indexes, and testing the contagion phenomenon versus the flight to quality (outflows of capital from stocks markets to bonds markets in crises periods). The copula introduced is a new variant of the Gumbel copula sufficiently flexible to describe asymmetric effects between the variables. This copula is designed to be capable of reflecting these asymmetric effects, and therefore to describe causality in the extreme values. The choice of the Gumbel copula is motivated by the multivariate extreme value theory and the dependence properties between the vector of maxima. The concepts of contagion and interdependence are revisited and adapted to be defined as tail properties.

Finally, the last chapter sketches the future lines of research of the author and the main conclusions found in the thesis.

Chapter 2

Risk is in the Tails: A formal definition of Extreme Values

The goal of this chapter is to define and estimate the extremes of any random sample of size n, from a distribution function F. This is done by means of a threshold sequence and a goodness of fit test. Extreme Values are the observations exceeding such threshold and following a type of Generalized Pareto distribution (GPD) involving the tail index of F. The threshold is the order statistic that minimizes a generalization of the distance of the supremum between the empirical distribution function of the corresponding largest observations and the corresponding GPD. This generalization includes the Kolmogorov-Smirnov statistics as a particular case. Once a set of extreme values candidates is identified, we use a semi-parametric bootstrap to test the corresponding GPD approximation (second part of our definition). Monte Carlo simulations show a very good finite sample performance of the proposed test. Finally, we use our methodology to estimate the tail index and Value at Risk of some financial indexes of major stock markets.

2.1 Introduction

Risk Management is one of the most important innovations of the 20th century in Economics. During the last decade financial markets have realized the importance of monitoring risk. The question one would like 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 $\{X_n\}$, 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 to know 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 have been recently introduced, like Value at Risk (VaR) or Expected Shortfall. These measures are functions of extreme quantiles of the data distribution. The attempt for modelling the tails of these distributions is troublesome and standard methodologies as historical simulation or the gaussian distribution do not provide reliable approximations at very high quantiles.

On the other hand, the methodology derived from EVT covers this gap, and produces a parametric framework to derive the VaR or any function of this extreme quantile. It is clear that the first task is to identify which values are really extreme values. In practice this is done by graphical methods like QQ-plot, Sample Mean Excess Plot or by other ad-hoc methods that impose an arbitrary threshold $(5\%, 10\%, \ldots)$, see Embrechts, Klüppelberg and Mikosch (1997). These methods do not propose any formal computable method, and moreover they only give very rough estimates of the set of extreme values. In this chapter we propose a formal way of identifying and estimating the extreme values of any random sample of size n coming from a distribution function, say F. These values are going to be defined as the exceedances of a threshold sequence $\{u_n\}$ following a type of Generalized Pareto distribution (GPD). The selection of this threshold plays a central role in this definition and in the estimation of the parameters of the GPD. The sequence of extreme values depends on the length of the data sequence by the choice of $\{u_n\}$. Therefore, we need to introduce an appropriate test to asses statistically whether the distribution function of the set of extremes candidates given by the threshold, really satisfies the weak convergence to the GPD or not, with parameters driven by F. In order to achieve this task, we propose a semi-parametric bootstrap test and study its asymptotic as well as its finite sample performance.

The final purpose of our methodology is to achieve a reliable approximation of F paying special attention to its tails. Our tail estimate provides accurate approximations of the extreme quantiles of F, and from them it is straightforward to calculate the risk measures introduced in the financial literature.

The chapter is structured as follows. In Section 2 we present some general results of extreme value theory focusing on the weak convergence of the largest observations of a random sequence. Section 3 introduces different approaches to select the threshold sequence and gives a brief review of estimation methods for the parameters of the GPD. Some simulations show the performance of our approach in terms of tail index estimation. The complete definition of the sequence of extreme values is given in section 4 by means of a bootstrap hypothesis test. Monte Carlo simulations provide the finite sample performance of our proposed test. Section 5 presents an empirical application where the risk of financial indexes of major stock markets is analyzed via the tail index and VaR. Finally, Section 6 draws some concluding remarks. Proofs are collected in the appendix.

2.2 Extreme Value Theory Results

The purpose of this section is to briefly introduce the set of results of the so called extreme value theory necessary to develop the theory used herein. The departing point is the study of the weak convergence for the sample maximum of a sequence of random variables $\{X_n\}$ with distribution function F. Our intention is to use the limiting distribution of this statistic to derive the weak convergence of the largest observations of a random sequence imposing a minimum set of assumptions on the distribution function F.

Let $M_n = max\{X_1, \ldots, X_n\}$ be the sample maximum of the sequence and let F be the common distribution function for $\{X_n\}$. Our first goal is to introduce the conditions under which M_n converges weakly to a non degenerate distribution function.

Result 2.2.1. *Let $\{X_n\}$ be an independent and identically distributed (iid) sequence. Let $0 \leq \tau \leq \infty$ and suppose that $\{u_n\}$ is a sequence of real numbers such that*

$$n(1 - F(u_n)) \rightarrow \tau \quad as \quad n \rightarrow \infty. \tag{2.1}$$

Then

$$P\{M_n \leq u_n\} \rightarrow e^{-\tau} \quad as \quad n \rightarrow \infty. \tag{2.2}$$

Conversely, if condition (2.2) *holds for some τ, $0 \leq \tau \leq \infty$, then so does condition* (2.1).

The proof of this result is immediately derived from

$$P\{M_n \leq u_n\} = F^n(u_n) = (1 - \frac{n(1 - F(u_n))}{n})^n. \tag{2.3}$$

However this result does not guarantee the existence of a non degenerate distribution for M_n. Define the right end point of a distribution function as $x_F = \sup\{x|F(x) < 1\} \leq +\infty$. It is clear that $M_n \rightarrow x_F$ with probability 1 as $n \rightarrow \infty$. Suppose now that F has a jump at x_F with $x_F < \infty$ (*i.e.* $F(x_{F^-}) < 1$), and consider a sequence $\{u_n\}$ satisfying (2.2) with $0 \leq \tau \leq \infty$. Then, either $u_n < x_F$ for infinitely many values of n and $n(1 - F(u_n)) \rightarrow \infty$, or $u_n > x_F$ and $n(1 - F(u_n)) = 0$. Therefore we also need some regularity condition on the tail of F to avoid the existence of jumps.

Result 2.2.2. *Let F be a distribution function with right end point x_F such that*

$$\lim_{x \uparrow x_F} \frac{1 - F(x)}{1 - F(x^-)} = 1, \tag{2.4}$$

and let $\{u_n\}$ be a sequence with $u_n < x_F$ and $n(1 - F(u_n)) \rightarrow \tau$. Then $0 < \tau < \infty$.

The choice of the sequence $\{u_n\}$ determines the value of τ. Suppose $v_n > u_n$ and (2.2) holds, then $n(1 - F(v_n)) \rightarrow \tau'$ with $\tau' < \tau$. We can write expression (2.2) as $P\{M_n \leq u_n(x)\} \rightarrow e^{-\tau(x)}$, with u_n depending on x. Moreover, there exist some scaling sequences a_n, b_n varying according to F such that

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x) \quad as \quad n \rightarrow \infty, \tag{2.5}$$

with $u_n(x) = a_n x + b_n$ and $G(x) = e^{-\tau(x)}$ a distribution function. This function has been fully characterized by Gnedenko (1943) or de Haan (1976) via the analysis of domains of attraction for the maximum, and it can be summarized as follows:

Result 2.2.3. *The distribution function $G(x)$ derived in expression* (2.5) *can only take three different forms,*

Type I: (Gumbel) $\quad G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty,$

Type II: (Frèchet) $\quad G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\frac{1}{\xi}}} & x > 0, \xi > 0 \end{cases}$

Type III: (Weibull) $\quad G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(-x)^{-\frac{1}{\xi}}} & x < 0, \xi < 0 \end{cases}.$

The parameter ξ is the tail index of F and characterizes the tail behavior of the distribution function. The three types can be gathered in the so called Generalised Extreme Value Distribution, first proposed by von Mises (1936),

$$G(x) = e^{-(1+\xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}}, \tag{2.6}$$

where μ is a location parameter, σ a scale parameter and $\xi \neq 0$. This expression boils down to $G(x) = e^{-e^{-\frac{x-\mu}{\sigma}}}$ when $\xi = 0$.

Clearly $\tau(x) = (1 + \xi\frac{x-\mu}{\sigma})^{\frac{-1}{\xi}}$ in expression (2.5), and hence $n(1 - F(u_n(x))) \to (1 + \xi\frac{x-\mu}{\sigma})^{\frac{-1}{\xi}}$ for all x, where a_n, b_n are suitable constants. This is the result we exploit in order to derive the weak convergence of the largest observations determined by a threshold sequence $u_{on} = a_n\mu + b_n$, with μ satisfying $-log\ G(\mu) = 1$. By doing that

$$\frac{1 - F(u_n(x))}{1 - F(u_{on})} \to (1 + \xi\frac{x - \mu}{\sigma})^{\frac{-1}{\xi}}, \quad as \quad n \to \infty. \tag{2.7}$$

This expression can be rewritten as

$$\frac{F(u_n(x)) - F(u_{on})}{1 - F(u_{on})} \to 1 - (1 + \xi\frac{x - \mu}{\sigma})^{-\frac{1}{\xi}}, \tag{2.8}$$

for all $x > \mu$ continuity point. The threshold sequence satisfies $u_n(x) = u_{on} + a_n(x - \mu)$, and we can define

$$F_{u_{on}}(a_n(x - \mu)) = \frac{F(u_{on} + a_n(x - \mu)) - F(u_{on})}{1 - F(u_{on})}, \tag{2.9}$$

as the conditional excess distribution function given u_{on} with $x > \mu$. This takes us directly to the following result:

Result 2.2.4. *Let $y = a_n(x - \mu)$, then*

$$\lim_{u_{on} \to x_F} \sup_{[0 \le y < \infty]} |F_{u_{on}}(y) - GPD_{\xi,\sigma(u_{on})}(y)| = 0, \tag{2.10}$$

with

$$GPD_{\xi,\sigma(u_{on})}(y) = \begin{cases} 1 - (1 + \xi\frac{y}{\sigma(u_{on})})^{-\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - e^{\frac{-y}{\sigma(u_{on})}} & if\ \xi = 0 \end{cases}, \tag{2.11}$$

the Generalized Pareto distribution and $\sigma(u_{on}) = \sigma a_n$.

This result is known as Pickands (1975) theorem. Pickands proposed a sequence u_{on} taken in the interval $[b_n, b_{n+1}]$ with b_n the suitable sequence in (2.5). This approximation for the distribution of the largest observations regarded as the exceedances of a threshold sequence can be improved when the tail of F decays at a polynomial rate.

Suppose $1 - F(x) = x^{-\frac{1}{\xi}} L(x)$ with $L(tx)/L(x) \to 1$ as $x \to x_F$ and $\xi > 0$; then the distribution function F satisfies

$$\lim_{x \uparrow x_F} \frac{1 - F(tx)}{1 - F(x)} = t^{-\frac{1}{\xi}}, \qquad t > 0. \tag{2.12}$$

This type of distribution functions are regularly varying at a rate $\frac{1}{\xi}$ and the domain of attraction of the sample maximum is the Fr$\acute{e}$chet distribution (see Resnick (1987) or de Haan (1976)). The function $L(x)$ is said slowly varying and is introduced to include the deviations of F from the Pareto probability law. When these departures from the polynomial law are small, $F_{u_{on}}(y)$ is better approximated by the Pareto distribution function. Consider a sequence $u_n(x) = u_{on}x$ where $u_{on} = u_n(1)$ is the threshold sequence that satisfies $1 - F(u_{on}) = u_{on}^{-\frac{1}{\xi}} L(u_{on})$. The conditional excess distribution function defined by u_{on} as $F_{u_{on}}(u_n(x)) = \frac{F(u_n(x)) - F(u_{on})}{1 - F(u_{on})}$ satisfies

$$F_{u_{on}}(u_n(x)) \to 1 - (\frac{u_n(x)}{u_{on}})^{-\frac{1}{\xi}}, \quad as \quad n \to \infty, \tag{2.13}$$

for $u_n(x) \geq u_{on}$ or equivalently for $x \geq 1$. This convergence holds for all continuity point of F and therefore for this case we can rewrite the previous result as

$$\lim_{u_{on} \to x_F} \sup_{[u_{on} \leq y < \infty]} |F_{u_{on}}(y) - PD_\xi(y)| = 0, \tag{2.14}$$

with $y = u_n(x)$ and $PD_\xi(y) = 1 - (\frac{y}{u_{on}})^{\frac{-1}{\xi}}$.

Finally, the choice of the threshold sequence has also an effect on the error made by the approximations claimed in Pickands theorem. This error arises from the asymptotic relation $n(1 - F(u_n)) \to \tau$ and from the approximation of $F^n(u_n)$ by the exponential distribution. The latter approximation is of order $o(n^{-1})$ since

$$0 \leq e^{-x} - (1 - \tfrac{x}{n})^n \leq 0.3 \tfrac{1}{n-1},$$

for $0 \leq x \leq n$ (see, e.g., Leadbetter, Lindgren and Rootzén (1983)). Nevertheless, if F is continuous one can always obtain an equality in expression (2.2) by taking $u_n = F^{-1}(e^{-\frac{\tau}{n}})$ and making the approximation errors to vanish. However sequences of type $u_n(x) = a_n x + b_n$, with a_n, b_n suitable constants are more appropriate to study the weak convergence of M_n. In these cases the equality or any uniform bound for all x are not usually feasible in expression (2.5).

2.3 Threshold Choices to define the Extreme Values

The last section has focused on finding the asymptotic laws that rule the largest observations of a random sequence from a distribution function F. This set of observations is defined by means of a threshold sequence and the tail index ξ that characterizes the corresponding Generalized Pareto or Pareto. The choice of this sequence is troublesome since $u_{on} \to x_F$ when $n \to \infty$, but at an appropriate rate. This order of convergence depends on F represented by the sequences a_n and b_n when $u_n(x)$ is of the form $u_n(x) = a_n x + b_n$. Hence the threshold sequence u_{on} can be defined by the scaling sequences a_n, b_n and the value of x satisfying the condition $-log\ G(x) = 1$, or equivalently $n(1 - F(u_{on})) \to 1$. For ease of notation we will use hereafter u_n instead of u_{on} to denote the threshold sequence satisfying these conditions. This sequence is immediately derived by direct calculations when F is known. Consider as an example the case $F(x) = 1 - e^{-x}$. By continuity of F we can choose $u_n(x) = F^{-1}(1 - \frac{\tau(x)}{n})$ with $\tau(x) > 0$, and hence $u_n(x) = -log\ \tau(x) + log\ n$. Expression (2.2) is written as

$$P\{M_n \leq -log\ \tau(x) + log\ n\} \to e^{-\tau(x)},$$

and then $P\{M_n - log\ n \leq x\} \to e^{-e^{-x}}$, with $\tau(x) = e^{-x}$ for all $x > 0$. The scaling constants are $a_n = 1$, $b_n = log\ n$, and hence the threshold sequence is $u_n = log\ n$, since $-log\ G(0) = 1$. More examples can be found in Leadbetter, Lindgren and Rootzén (1983).

In general F is unknown, and in this setting neither the theoretical derivation nor the direct comparison of different threshold choices are possible. This comparison is undertaken by analyzing the properties of the tail index estimator of F, as most of these estimators for ξ are tied to a threshold choice. Therefore their biases and variances are influenced by the effect of the selection of u_n. There is a large amount of literature in tail index estimation (chapter VI of Embrechts, Klüppelberg and Mikosch (1997) gives an excellent review). Among these estimators, the most popular are Hill's estimator (1975), and Pickands's estimator (1975). The former is given by

$$\hat{\xi}_n^{Hi}(u_n) = \frac{1}{k} \sum_{i=n-k+1}^{n} \log \frac{x_{(i)}}{x_{(n-k)}}, \tag{2.15}$$

with $u_n = x_{(n-k)}$, $x_{(n-k+1)} \leq \ldots \leq x_{(n)}$ denoting the increasing order statistics and k an integer value in $[1, n]$. Pickands's estimator for the tail index is

$$\hat{\xi}_n^{Pi}(u_n) = \frac{1}{\log(2)} \log\left(\frac{x_{(n-k+1)} - x_{(n-2k+1)}}{x_{(n-2k+1)} - x_{(n-4k+1)}}\right), \tag{2.16}$$

and

$$\hat{\sigma}_n^{Pi}(u_n) = \frac{x_{(n-2k+1)} - x_{(n-4k+1)}}{\int_0^{\log 2} e^{\hat{\xi}_n^{Pi}(x_{(n-4k+1)})t} dt}, \tag{2.17}$$

for the variance, with $u_n = x_{(n-4k+1)}$ and $k = 1, \ldots, n/4$. There are some features of both estimators that is worth mentioning. These estimators are heavily dependent on the threshold choice u_n, and both of them can be derived under the assumption that F_{u_n} is exactly Pareto with parameter ξ or Generalized Pareto with parameters ξ and $\sigma(u_n)$. Moreover, if $F_{u_n} = PD_\xi$, Hill's estimator is the maximum likelihood estimator (Ml) of ξ inheriting the corresponding asymptotic properties: consistency and normal distribution. This approach is only valid for regularly varying distribution functions, $i.e.$ $\xi > 0$, otherwise the asymptotic properties of this estimator vary according to F (see Davis and Resnick (1984)).

Pickands's estimator for the tail index is obtained assuming $F_{u_n} = GPD_{\xi,\sigma(u_n)}$ and taking the inverse of the parametric GPD. This estimator is consistent and also converges to a normal distribution; but is very sensitive to the choice of u_n. Alternatively, under the latter parametric assumption on F_{u_n} we can obtain the maximum likelihood estimator for the parameter ξ and $\sigma(u_n)$ of the GPD. In this case there is not a closed expression for the maximum likelihood estimators of these parameters, and we have to rely on numerical procedures (see Press (1992)). The maximum likelihood estimator for the tail index is consistent and asymptotically normal for $\xi > -\frac{1}{2}$, as it is discussed in Smith (1985).

The threshold selection is carried out by studying the mean square error of these ξ estimators, as u_n is let to vary. However some explicit form is required for the distribution function F. Under the assumption

$$1 - F(x) = Cx^{-\frac{1}{\xi}}[1 + Dx^{-\beta} + o(x^{-\beta})], \tag{2.18}$$

where $\xi > 0, C > 0, \beta > 0$ and D is a real number, Hall (1982) proposed estimators for the tail index based on an optimal choice of intermediate order statistics as candidates for the threshold sequence. Nevertheless the pioneer work for threshold selection is Pickands (1975), where F is not necessarily as in (2.18). The estimation of the tail index and the threshold selection are done in one single step. Pickands proposed as a candidate for the threshold the order statistic of a sample $\{x_n\}$ that minimizes the distance d^∞ involving the distribution functions $F_{u_n,n}$ and $GPD_{\hat{\xi}_n^{Pi}(u_n),\hat{\sigma}_n^{Pi}(u_n)}$. The empirical conditional excess distribution function $F_{u_n,n}(x)$ with $x > u_n$ is defined by

$$F_{u_n,n}(x) = \sum_{i=1}^{n} \frac{1_{\{u_n < x_i \leq x\}}}{\sum_{j=1}^{n} 1_{\{x_j > u_n\}}}, \tag{2.19}$$

or equivalently, via the transformation $y = a_n(x - u_n) > 0$, by

$$F_{u_n,n}(y) = \sum_{i=1}^{n} \frac{1_{\{0 < y_i \leq y\}}}{\sum_{j=1}^{n} 1_{\{y_j > 0\}}}. \tag{2.20}$$

The distance d^∞ can be written as function of a parameter θ,

$$d^\infty(F_{\theta,n}, GPD_{\hat{\xi}_n^{Pi}(\theta),\hat{\sigma}_n^{Pi}(\theta)}) = \sup_{0 \leq y < \infty} |F_{\theta,n}(y) - GPD_{\hat{\xi}_n^{Pi}(\theta),\hat{\sigma}_n^{Pi}(\theta)}(y)|. \tag{2.21}$$

The optimal threshold is then

$$u_n^{Pi} = \arg\min_{\theta} \quad d^{\infty}(F_{\theta,n}, GPD_{\hat{\xi}_n^{Pi}(\theta),\hat{\sigma}_n^{Pi}(\theta)}), \tag{2.22}$$

with θ taking values along the ordered sample $x_{(3n/4)} \leq \ldots \leq x_{(n)}$. More specifically, $u_n^{Pi} = x_{(n-k)}$ with $k \to \infty$, $n \to \infty$ and $k = o(n)$ to benefit of an increase in the sample size.

Alternatively, we propose a version of the distance d^{∞} where the number of tail observations is weighted differently. This new approach accounts for the estimation pitfalls that derive from the lack of observations when θ gets close to x_F.

Definition 2.3.1. *Let $F_{\theta,n}$ be the empirical version of F_θ and $GPD_{\hat{\xi}_n^{Ml}(\theta),\hat{\sigma}_n^{Ml}(\theta)}$ the distribution function of the largest observations with parameters estimated by maximum likelihood (Ml). Define the Weighted Pickands distance d^{Wp} as*

$$d^{Wp}(F_{\theta,n}, GPD_{\hat{\xi}_n^{Ml}(\theta),\hat{\sigma}_n^{Ml}(\theta)}) = k^{\varepsilon} \sup_{0 \leq y < \infty} |F_{\theta,n}(y) - GPD_{\hat{\xi}_n^{Ml}(\theta),\hat{\sigma}_n^{Ml}(\theta)}(y)|, \tag{2.23}$$

with $0 \leq \varepsilon \leq 1/2$ and $k = \sum_{j=1}^{n} 1_{\{x_j > \theta\}}$.

The parameter ε determines the weight assigned by the distance d^{Wp} to the tail observations defined by the corresponding parameter θ. Notice that this distance is the one used by Pickands when $\varepsilon = 0$, and the Kolmogorov-Smirnov (KS) statistic (Kolmogorov (1933)) when $\varepsilon = 1/2$. The corresponding threshold choice is the order statistic that minimizes the distance,

$$u_n = \arg\min_{\theta} \quad d^{Wp}(F_{\theta,n}, GPD_{\hat{\xi}_n^{Ml}(\theta),\hat{\sigma}_n^{Ml}(\theta)}), \tag{2.24}$$

with θ taking values along the ordered sample $x_{(1)} \leq \ldots \leq x_{(n)}$. The parameter ε can be useful to study the effect of different weighting schemes in the threshold selection; however this is far beyond the scope of this chapter where we will only focus on the value $\varepsilon = 1/2$ (KS statistic).

A preliminary analysis points out that threshold values far from x_F produce biased estimates of the tail index. On the other hand, u_n close to the right end point will result in inefficient estimates of ξ.

Goldie and Smith (1987) and Smith (1987) derive the asymptotic distribution functions of both the *Ml* and Hill estimators of the tail index for a class of distribution functions such that $1 - F(x) = x^{-\frac{1}{\xi}} L(x)$, where $L(x)$ are slowly varying functions of different types. They also discuss in detail asymptotic bias and variance for these estimators and find that departures of F from a Pareto distribution function lead to biased and inefficient estimates of the tail index for both estimators. As a result, a right choice of the threshold sequence turns out to be of critical importance in order to minimize the mean square error (MSE).

Hall (1982) derives an analytical expression for the MSE of Hill's estimator when F satisfies (2.18). All these results are achieved for determined classes of distribution functions. In contrast, under the set of assumptions stated in the previous section it is not possible to derive analytically the mean square error expression for the tail index estimator. Therefore, we propose bootstrap confidence intervals in order to measure the bias and uncertainty of the different tail index estimators we considered.

The naïve nonparametric bootstrap is consistent since the empirical distribution function F_n is a consistent estimator of F and $\sqrt{k}(\hat{\xi}_n^{(i)}(u_n) - \xi)$, $i = Hi, Ml$ converges weakly to a normal distribution, with k the number of exceedances over u_n. Then, the bootstrap approximation $J_n(x, F_n)$ to the true sampling distribution function $J_n(x, F)$ of this statistic can be used to produce confidence regions, at $1 - \alpha$ level, in the following way

$$\xi \in [\hat{\xi}_n^{Ml}(u_n) - \frac{1}{\sqrt{k}} J_n^{-1}(1 - \frac{\alpha}{2}, F_n), \hat{\xi}_n^{Ml}(u_n) - \frac{1}{\sqrt{k}} J_n^{-1}(\frac{\alpha}{2}, F_n)], \tag{2.25}$$

where $J_n^{-1}(1 - \alpha, F_n)$ is the $1 - \alpha$ bootstrap quantile. To implement (2.25) the bootstrap approximation is estimated by

$$\hat{J}_n(x, F_n) = \frac{1}{B} \sum_{j=1}^{B} 1_{\{\sqrt{k}(\hat{\xi}_{j,n}^{*Ml}(u_{j,n}^{*}) - \hat{\xi}_n^{Ml}(u_n)) \leq x\}}, \tag{2.26}$$

with B the number of bootstrap iterations, $\hat{\xi}^{*Ml}_{j,n}(u^*_{j,n})$ the maximum likelihood estimator for the bootstrap sample j, and $u^*_{j,n}$ the corresponding threshold choice.

The finite sample performance of the different estimators is analyzed in Table 2.8.1. The threshold u_n is chosen by both methods, Pickands and Weighted Pickands with $\epsilon = 1/2$. To remark the importance of the threshold selection to estimating the tail index, an ad-hoc threshold ($u_n = x_{(\frac{95}{100}n)}$) is also included in the analysis.

The simulation experiment of Table 2.8.1 is done for different t-student distributions, where the tail index ξ is well approximated by the inverse of the degrees of freedom (see chapter *III* of Embrechts, Klüppelberg and Mikosch (1997)).

Before discussing the results of Table 2.8.1 it is important to notice that although F is known, we replace it by F_n to calculate the bootstrap approximation $J_n(x, F_n)$. The reason to do that is that this bootstrap procedure works even when F is unknown and we only have a realization from the random sequence $\{X_n\}$.

There are two clear results from Table 2.8.1: First, the confidence intervals for our estimator contain the true tail index, something that it does not occur for Pickands's method; and second the confidence intervals estimated from the ad-hoc threshold are wider than the ones derived from our method when ξ is significantly greater than zero.

Table 2.8.2 analyzes more in detail the advantages of the Weighted Pickands method for selecting u_n when the data come from heavy tailed distributions. In this case the $GPD_{\xi,\sigma(u_n)}$ is replaced by the PD_ξ in (2.3.1) and (2.24).

From Table 2.8.2 we conclude that when we are dealing with heavy tailed distributions ($\xi > 0$), our method is more efficient with the PD rather than with GPD. These simulation results are in the same line as the theoretical findings derived in Smith (1987).

2.4 Hypothesis Testing

Different threshold choices define different sets of possible extreme values of a particular sequence $\{X_n\}$. In this chapter the observations exceeding certain threshold are considered extreme values only if additionally they are distributed as a $GPD_{\xi,\sigma(u_n)}$, with ξ the tail index of F. In order to check this condition we propose a goodness of fit test for the following hypothesis:

$$H_{n,0}\text{: the sample } \{(x_1 - u_n)_+, \ldots, (x_n - u_n)_+\} \text{ is distributed as } GPD_{\xi,\sigma(u_n)}$$

versus a general alternative of the form

$$H_{n,1}\text{: the sample } \{(x_1 - u_n)_+, \ldots, (x_n - u_n)_+\} \text{ is not distributed as } GPD_{\xi,\sigma(u_n)}$$

with $u_n \in \mathbb{R}$, ξ the tail index of F and $(x)_+ = max(x, 0)$.

A natural goodness of fit test statistic is the KS statistic (for other goodness of fit criteria see Anderson and Darling (1952)),

$$R_k(y; \xi, \sigma(u_n)) = \sqrt{k} \sup_{0 \le y < \infty} |P_k(y) - GPD_{\xi,\sigma(u_n)}(y)|, \tag{2.27}$$

with $k = \sum_{j=1}^{n} 1_{\{x_j > u_n\}}$ and P_k the empirical distribution function of the observations exceeding u_n. When the parameters are known, the asymptotic distribution of this test statistic is tabulated and the critical values can be derived. If the parameters are unknown but consistently estimated, the bootstrap distribution function is a reliable approximation of the true sampling distribution of $R_k(y; \xi, \sigma(u_n))$. In this case it can be proved (see Romano (1988)) that the bootstrap critical values are consistent estimates of the actual ones.

Our interest, however, does not lie in the definition of the extreme values of a particular sequence $\{X_n\}$; but in the definition of the extreme values of any sequence of length n with distribution function F. In this case a different hypothesis test is needed to determine whether the selected threshold is a good candidate to define the extremes of F given the sample size n. More formally, the testing problem under consideration is

$$H_0 : F_{u_n} = GPD_{\xi,\sigma(u_n)}$$

versus a general alternative

$$H_1 : F_{u_n} \neq GPD_{\xi,\sigma(u_n)},$$

with ξ being the tail index of F.

Now, we can formally define the set of extreme values of any sequence with distribution function F.

Definition 2.4.1. *Let $\{X_n\}$ be any sequence of a distribution function F. The extreme values of any sequence of length n from this distribution are given by the observations exceeding the threshold u_n, and satisfying $F_{u_n} = GPD_{\xi,\sigma(u_n)}$.*

The test statistic in this case is a version of the family of KS test statistics,

$$T_n(y_n; \xi, \sigma(u_n)) = \sqrt{n} \sup_{0 \leq y < \infty} |F_{u_n,n}(y) - GPD_{\xi,\sigma(u_n)}(y)|, \tag{2.28}$$

with $y_i = (x_i - u_n)_+$, $i = 1, \ldots, n$. This statistic depends on u_n, ξ and $\sigma(u_n)$.

In order to derive the asymptotic distribution of (2.28) and to asses the bootstrap approximation, the following results are required. Let

$$U_\lambda(t) = \frac{P\{\lambda < T \leq t\}}{P\{T > \lambda\}} \tag{2.29}$$

be the conditional excess distribution function, with parameter λ on $[0, 1]$, of a uniform $[0, 1]$ random variable T. Its empirical counterpart

$$U_{\lambda,n}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{\lambda < t_i \leq t\}}}{\frac{1}{n}\sum_{j=1}^{n} 1_{\{t_j > \lambda\}}}, \tag{2.30}$$

with $t_1, \ldots, t_n$ and $t \in [0, 1]$, defines an empirical process $B_n(t) = \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))$ similar to the uniform empirical process $\sqrt{n}(U_n(t) - U(t))$. It is well known that the latter converges weakly to the distribution of a mean zero gaussian process $Z_U(\cdot)$ (see chapter V of Pollard (1984)). By an analogue reasoning, it is immediate to derive the probability law of the process $S_n(y) = \sqrt{n}(F_{u_n,n}(y) - F_{u_n}(y))$ where the threshold u_n plays the role of the parameter λ.

Theorem 2.4.1. *Consider a continuous and strictly increasing distribution function F and a threshold u_n, with $u_n < x_F$. The empirical process $S_n(y)$ converges weakly to the distribution of a mean zero gaussian process $Z_{F_{u_n}}(\cdot)$ with covariance function*

$$Cov(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{(F(min(y_1,y_2)) - F(u_n)) - (F(y_1) - F(u_n))(F(y_2) - F(u_n))}{(1-F(u_n))^2}, \tag{2.31}$$

with $y_1, y_2 \in \mathbb{R}$. Moreover, under the null hypothesis H_0, this empirical process takes the form $\sqrt{n}(F_{u_n,n}(y) - GPD_{\xi,\sigma(u_n)}(y))$ and the covariance function becomes

$$Cov(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{GPD_{\xi,\sigma(u)}(min(y_1,y_2))}{1-F(u_n)} - GPD_{\xi,\sigma(u_n)}(y_1)GPD_{\xi,\sigma(u_n)}(y_2). \tag{2.32}$$

By the continuous mapping theorem, the limiting distribution function, denoted by $L(x,F)$, of the test statistic T_n is the distribution of the supremum of a mean zero gaussian process with covariance function (2.32). The proof is in the appendix.

In order to test H_0, we should be using the following rejection criteria

$$\{T_n(y_n; \xi, \sigma(u_n)) > L_n^{-1}(1-\alpha, F)\}, \tag{2.33}$$

where $L_n^{-1}(1-\alpha, F)$ is the $1-\alpha$ quantile of the exact finite sample distribution $L_n(x,F)$ of the statistic T_n. This distribution L_n is clearly unknown and in practice has to be approximated by the asymptotic distribution $L(x,F)$. This limiting distribution takes a complicated form and depends on the knowledge of F, on the parameters of the GPD, as well as on the threshold u_n. The nuisance parameters dependency forces us to look for an alternative method to approximate the distribution $L_n(x,F)$.

2.4.1 Bootstrap Approximation

Let $L_n(x, Q_n)$ be the bootstrap distribution that approximates $L_n(x, F)$, and $L_n^{-1}(1-\alpha, Q_n)$ the bootstrap quantile that approximates the corresponding finite sample distribution quantile $L_n^{-1}(1-\alpha, F)$. In order for the bootstrap to be consistent, Q_n has to satisfy certain conditions.

Lemma 2.4.1. *Let Q_n be an estimator of F based on $\{x_1, \ldots, x_n\}$ that satisfies $\sup_{x\in\mathbb{R}} |Q_n(x) - F(x)| \xrightarrow{p} 0$ whenever $F \in H_0$, and let $L(x, F)$, the limiting distribution of the test statistic T_n, be continuous and strictly increasing. Then*

$$P\{T_n > L_n^{-1}(1-\alpha, Q_n)\} \to \alpha, \quad as \quad n \to \infty. \tag{2.34}$$

The naïve nonparametric bootstrap from $Q_n = F_n$ fails to produce consistent estimates of a distribution function under H_0 if F does not belong to the null. On the other hand, the parametric bootstrap from the $GPD_{\xi,\sigma(u_n)}$ (see (2.27)) fails to capture the structure of F for the observations smaller than the threshold u_n.

To fulfill the conditions of Lemma (2.4.1) corresponding to Q_n and therefore to solve the two previously mentioned problems, a semi-parametric bootstrap methodology is introduced. Define

$$Q_n(x) = \begin{cases} F_n(x) & x \le u_n \\ GPD_{\xi,\sigma(u_n)}(x-u_n) + F_n(u_n)(1 - GPD_{\xi,\sigma(u_n)}(x-u_n)) & x > u_n. \end{cases} \tag{2.35}$$

This distribution function is derived from the conditional probability theorem, since

$$P\{X \le x\} = P\{X \le u_n\}P\{X \le x \mid X \le u_n\} + P\{X > u_n\}P\{X \le x \mid X > u_n\}, \tag{2.36}$$

where $P\{X \le u_n\}$ is consistently approximated by $F_n(u_n)$, and under the null $P\{X \le x \mid X > u_n\} = GPD_{\xi,\sigma(u_n)}(y)$ with $y = x - u_n$.

Denote $\{x_n^*\}$ a bootstrap sample obtained from Q_n and consider the transformed bootstrap sample $y_i^* = x_i^* - u_n$ with $i = 1, \ldots, n$. The value of the test statistic is $t_n(y_1^*, \ldots, y_n^*; \xi, \sigma(u_n))$ and for the sake of notation is denoted as $t_n^*(y_n; \xi, \sigma(u_n))$. The bootstrap approximation $L_n(x, Q_n)$ is then estimated by the empirical distribution of the B (number of bootstrap samples) values of T_n,

$$\hat{L}_n(x, Q_n) = \frac{1}{B}\sum_{j=1}^{B} 1_{\{t_{n,j}^*(y_n;\xi,\sigma(u_n)) \leq x\}}. \tag{2.37}$$

The $1-\alpha$ quantile of $\hat{L}_n(x, Q_n)$ is the order statistic $t^*_{n,(\lceil (1-\alpha)B \rceil)}(y_n; \xi, \sigma(u_n))$ of the sequence $\{t_{n,j}^*(y_n; \xi, \sigma(u_n))\}$ of B elements, where $\lceil x \rceil$ is the upper integer part of x. The rejection criteria (2.33) is replaced now by

$$\{T_n(y_n; \xi, \sigma(u_n)) > t^*_{n,(\lceil (1-\alpha)B \rceil)}(y_n; \xi, \sigma(u_n))\}, \tag{2.38}$$

and hence for a sample $\{x_n\}$ the null hypothesis is rejected if $t_n(y_1, \ldots, y_n; \xi, \sigma(u_n))$ is in this rejection region. This means that the conditional excess distribution function defined by u_n is not a $GPD_{\xi,\sigma(u_n)}$, and according to our definition these candidates for extreme observations are not really extreme.

Recall that until now we have assumed the parameters to be known. Nevertheless this condition is rarely satisfied in practice. To make our test operational we replace these parameters by their maximum likelihood estimators, and instead of Q_n we define its counterpart distribution function $\hat{Q}_n$:

$$\hat{Q}_n(x) = \begin{cases} F_n(x) & x \leq u_n \\ GPD_{\hat{\xi}_n^{Ml}(u_n),\hat{\sigma}_n^{Ml}(u_n)}(x-u_n) + F_n(u_n)(1 - GPD_{\hat{\xi}_n^{Ml}(u_n),\hat{\sigma}_n^{Ml}(u_n)}(x-u_n)) & x > u_n \end{cases}. \tag{2.39}$$

Notice that the new bootstrap distribution function $L_n(x, \hat{Q}_n)$ boils down to $L_n(x, Q_n)$ for $x \leq u_n$, and for $x > u_n$ the former $\sqrt{k}$-converges to the latter where k is the number of observations of the tail defined by u_n. Moreover, if F belongs to the null hypothesis defined by u_n, the conditions in Lemma (2.4.1) still hold and the rejection region (2.38) becomes

$$\{\hat{T}_n(y_n;\hat{\xi}_n^{Ml}(u_n),\hat{\sigma}_n^{Ml}(u_n)) > t^*_{n,(\lceil(1-\alpha)B\rceil)}(y_n;\hat{\xi}_n^{*,Ml}(u_n),\hat{\sigma}_n^{*,Ml}(u_n))\}, \tag{2.40}$$

where $\hat{T}_n$ and $\hat{\xi}_n^{Ml}(u_n)$, $\hat{\sigma}_n^{Ml}(u_n)$ are calculated from the original sample $\{x_n\}$ and $\hat{\xi}_n^{*,Ml}(u_n)$, $\hat{\sigma}_n^{*,Ml}(u_n)$ are estimated from the corresponding bootstrap sequences.

2.4.2 Finite Sample Performance: Empirical Power

The power of our test,

$$P\{\hat{T}_n > L_n^{-1}(1-\alpha,\hat{Q}_n)\}, \tag{2.41}$$

depends on three key parameters: the threshold choice, the distribution function F and the length of the sequence. To calculate this power is important to realize that the maximum likelihood estimates $\hat{\xi}_n^{Ml}(u_n),\hat{\sigma}_n^{Ml}(u_n)$ that entry in the expression of $\hat{T}_n$ are the ones used to define the null distribution $\hat{Q}_n$.

This test lies in constructing a distribution function $\hat{Q}_n$, such that its conditional excess distribution is a $GPD_{\hat{\xi}_n^{Ml}(u_n),\hat{\sigma}_n^{Ml}(u_n)}$. In that way the observations coming from the null hypothesis are drawn from $\hat{Q}_n$ and not from F. The empirical size of the test is calculated from the former distribution. For a deeper insight of how to calculate the power via bootstrap, see Beran (1986) and Romano (1988).

The following algorithms are devoted to describe the simulation experiment. Algorithm 2.4.1 generates bootstrap samples $\{x_n^*\}$ from the distribution function $\hat{Q}_n$ and calculates the empirical bootstrap approximation of $L_n(x,F)$. The threshold value u_n and the maximum likelihood estimates are obtained from a particular sample $\{x_n\}$ from F, and are used to construct $\hat{Q}_n$.

Algorithm 2.4.1. ***(Bootstrap procedure):***

1. *$l=1$.*

2. *Generate $x^*_{1,l},\ldots,x^*_{n,l}$ drawn from $\hat{Q}_n$.*

3. *Calculate* $\hat{\xi}_n^{*Ml}(u_n)$ *and* $\hat{\sigma}_n^{*Ml}(u_n)$ *from the bootstrap sample.*

4. $t_{n,l}^*(y_n;\hat{\xi}_n^{*Ml}(u_n),\hat{\sigma}_n^{*Ml}(u_n)) = \sqrt{n} \sup_{0\le y<\infty} |F_{u_n,n}(y) - GPD_{\hat{\xi}_n^{*Ml}(u_n),\hat{\sigma}_n^{*Ml}(u_n)}(y)|$ *with* $y = x - u_n$.

5. $l++$. *Go to step* 2 *while* $l \le B$.

6. $\hat{L}_n(x,\hat{Q}_n) = \frac{1}{B}\sum_{j=1}^{B} 1_{\{t_{n,j}^*(y;\hat{\xi}_n^{*Ml}(u_n),\hat{\sigma}_n^{*Ml}(u_n))\le x\}}$

In practice, the p-value replaces the rejection criteria given in expression (2.40). The empirical p-value is

$$p = \frac{1}{B}\sum_{j=1}^{B} 1_{\{t_{n,j}^*>\hat{t}_n\}}, \tag{2.42}$$

with $\hat{t}_n$ obtained from the sample $\{x_n\}$.

The probability (2.41) can not be directly derived and we have to rely on Monte Carlo simulations to calculate it. The following algorithm describes how to implement this procedure.

Algorithm 2.4.2. ***(Empirical Power):***

1. $j = 1$.

2. *Let* $\{x_{1,j},\dots,x_{n,j}\}$ *be a sample from* F *and obtain* u_n, $\hat{\xi}_n^{Ml}(u_n)$ *and* $\hat{\sigma}_n^{Ml}(u_n)$.

3. *Construct* $\hat{Q}_n$ *and* $\hat{L}_n(x,\hat{Q}_n)$ *as in algorithm* **??**.

4. *Generate* $\{x_1',\dots,x_n'\}$ *from* F_1.

5. *Calculate* $\hat{t}_n(x_n';\hat{\xi}_n^{Ml}(u_n),\hat{\sigma}_n^{Ml}(u_n))$ *if* $F_1 \neq F$. *Otherwise* $\hat{t}_n(x_n';\hat{\xi}_n^{*Ml}(u_n),\hat{\sigma}_n^{*Ml}(u_n))$ *with* $\hat{\xi}_n^{*Ml}(u_n),\hat{\sigma}_n^{*Ml}(u_n)$ *from* $\{x_n'\}$.

6. *Calculate the p-value* p *as in* (2.42).

7. $\delta_j = \begin{cases} 1 & \textit{if } p < \alpha \\ 0 & \textit{otherwise.} \end{cases}$

8. $j++$. *Repeat while* $j \leq m$.

9. $\hat{\alpha} = \frac{1}{m}\sum_{j=1}^{m}\delta_j$.

As $n \to \infty$, the estimate $\hat{\alpha}$ approaches the size of the test if the threshold u_n is really defining the extremes of F for a given length n. On the other hand, when the conditional distribution function defined by the threshold is not a $GPD_{\xi,\sigma(u_n)}$, or the sequence of data does not come from F the estimate $\hat{\alpha}$ tends to one.

The following Table 2.8.3 depicts the simulation results of the empirical power for a family of t-student distribution functions with the threshold u_n obtained by our Weighted Pickands Method.

Table 2.8.3 points out two clear results. First, the fact that the diagonal is very close to the nominal size reveals that our procedure performs very well to capture the extremes of sequences of length n coming from F_0. Second, extreme values candidates coming from $F_1 \neq F_0$ are rejected as extreme values of F_0. A by-product of this table is that our test can be considered a goodness of fit test via the tails. In principle our test is more sensitive than standard KS statistics to detect deviations in the tails (see Mason and Schuenemeyer (1983)).

Another alternative to select the threshold is to choose a fixed order statistic. In this case the set of extreme values is defined by a fixed number of observations given the sample size n.

The message from Table 2.8.4 is clear: These ad-hoc selections of the set of extreme values can be valid for particular sequences of F; but in general are rejected to define the extremes of any sequence of F with the same length n.

2.5 Application: VaR Estimation in Financial Indexes

An important application of the semi-parametric approximation $\hat{Q}_n$ of F is quantile estimation in the tail region, where there is usually a lack of observations because we are dealing with extremal events. This question is turning of primary importance in a wide variety of research fields, *e.g.* finance, climatology or hydrology.

The goal of this section is to get a deeper insight into risk management for financial indexes of different major markets. Market risk management is inherently related to the probability of occurrence of extreme events, that is, very large negative or positive returns. We focus on a particular measure of this market risk: Value at Risk (VaR), the amount of money necessary to provide the institution with coverage against losses that can occur with a p probability over some holding period. It is not our intention to get into details of the *VaR* methodology, we only pursue to present some results about tail index estimation (tail behavior) and a naïve calculus of *VaR* under *iid* assumptions for financial data. Of course we know this assumption is unrealistic and we should go a step further regarding heteroscedastic conditional volatility models; but this is left for future research.

General practitioners calculate VaRs in two different ways: (i) Complete parametric way, where it is assumed an underlying distribution (normal, t-student, etc.), and (ii) fully nonparametric way, where the main actor is the empirical distribution F_n. Our approach can be considered as something in the middle, because we use a semi-parametric approximation $\hat{Q}_n$.

The inverse of $\hat{Q}_n$ provides a consistent estimator of VaR for the distribution function F. In this case,

$$\widehat{\mathrm{VaR}}_p = \begin{cases} inf\{x|F_n(x) \geq 1-p\}, & 1-p \leq F_n(u_n) \\ u_n + \frac{\hat{\sigma}_n^{Ml}(u_n)}{\hat{\xi}_n^{Ml}(u_n)}((\frac{p}{1-F_n(u_n)})^{-\hat{\xi}_n^{Ml}(u_n)} - 1), & 1-p > F_n(u_n) \end{cases} . \tag{2.43}$$

When the distribution function is regularly varying ($\xi > 0$), the tail of $\hat{Q}_n$ is modelled as a Pareto distribution and the inverse of F is consistently estimated by

$$\widehat{\mathrm{VaR}}_p = \begin{cases} inf\{x|F_n(x) \geq 1-p\}, & 1-p \leq F_n(u_n) \\ u_n(\frac{1-F_n(u_n)}{p})^{\hat{\xi}_n^{Ml}(u_n)}, & 1-p > F_n(u_n) \end{cases} . \tag{2.44}$$

The uncertainty of these estimates can be measured by bootstrap confidence intervals, since the exact finite sample distribution function of $V_n = \sqrt{n}(\widehat{VaR_p} - VaR_p)$ is not known and its asymptotic distribution depends on nuisance parameters. Let $J_n(x, \hat{Q}_n)$ be the bootstrap approximation of the exact distribution of V_n. A confidence interval for VaR_p, at a significance level α, is therefore given by the following expression

$$I.C_\alpha(VaR_p) = [\widehat{\text{VaR}}_p - \frac{1}{\sqrt{n}} J_n^{-1}(1 - \frac{\alpha}{2}, \hat{Q}_n), \widehat{\text{VaR}}_p - \frac{1}{\sqrt{n}} J_n^{-1}(\frac{\alpha}{2}, \hat{Q}_n)], \tag{2.45}$$

where $J_n^{-1}(1 - \alpha, \hat{Q}_n)$ is the $1 - \alpha$ bootstrap quantile.

2.5.1 Data features

The data we use to illustrate how the methodology proposed in this work can be applied consist of five financial indexes of major stock markets over the period $19/12/1994 - 20/04/2001$: Frankfurt (DaX), London (Ftse-100), Madrid (Ibex), Tokyo (Nikkei) and New York (Dow-Jones). These data have been collected from *www.freelunch.com*. The observations considered for the analysis are the logarithmic returns measured in percentage terms and denoted as r_t:

$$r_t = 100 \ (logP_t - logP_{t-1}),$$

with P_t the original prices at time t. For ease of calculus the negative observations (losses) are depicted in the positive tail.

A first glance to the standard statistic for kurtosis shows that most of these series are leptokurtic. For instance the Dax index has a coefficient of corrected kurtosis of 5.70; Ftse: 1.34; Ibex: 3.88; Nikkei: 2.77, and the Dow-Jones has a coefficient of 3.25. Traditionally, this measure has been considered an indicator of heavy tails. Nevertheless, the coefficient of kurtosis does not provide us with adequate information about the source of the heaviness. The tail index, however, provides this kind of information focusing on a particular tail. For instance, $\xi > 0$ corresponds to distributions where that tail has a polynomial decay (a more detailed discussion can be found in Shiryaev (2001)).

Table 2.8.5 presents nonparametric bootstrap confidence intervals for the tail index (see (2.25)) obtained by the different approaches investigated throughout the work.

From this table appears clearly that the tail index ξ is greater than zero, indicating the existence of heavy right hand side tails (corresponding to losses). The only exception is $Ftse$ index, where there are some reasonable doubts. For that reason in the next table the VaR is calculated under both GPD and PD methodologies.

In Table 2.8.6 we provide pointwise estimates and confidence intervals for VaR under four different approaches. The first two correspond to the methods developed in this work, and the last two correspond to the standard empirical methodologies that will be used here as a benchmark.

From Table 2.8.6 three conclusions can be obtained: (i) Comparing our two approaches and taking into account the results of the previous table, the PD method outperforms the GPD from an efficiency point of view, given that the point estimates are very similar. This is the expected result under the presence of heavy tails; (ii) the approach based on the empirical distribution is less efficient, compared to the PD method. The main reason for that is the lack of observations coming from the tail, something that our PD method overcomes by parameterizing properly the tail; and (iii) the approach based on gaussianity, as expected, is very conservative in the sense of requiring less amount of capital (smaller VaR).

2.6 Conclusion

Risk and uncertainty are not the same thing (see Granger (2002)) and therefore they need to be characterized by different measures. It is accepted that variance is well designed to capture the latter but not the former. To measure risk, in other words, to respond the question *if things go wrong how wrong they can go?* it is first necessary to define the extreme observations that determine the risk underlying the sequence of data. This is the main goal of this work, where following Pickands (1975) methodology we do not only define formally the set of extreme observations of a particular sequence, but also, by means of a hypothesis test we define the extreme values of any sequence of the same length and with the same distribution function. Identification of the extreme observations allows to estimate very accurately risk measures such as Value at Risk, as well as to make inference on different tail parameters of interest.

The transmission of risk in time series, involving dependence in the largest observations, is developed in the following chapter.

2.7 Appendix A: Proofs

Proof of theorem 2.4.1: Let $\{U_n\}$ be a sequence of independent and identically distributed (iid) uniform random variables on $[0,1]$ and let λ be a parameter in $0<\lambda<1$. Define the empirical process $B_n(t) = \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))$ with $U_{\lambda,n}(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1_{\{\lambda<t_i\leq t\}}}{\frac{1}{n}\sum_{j=1}^{n}1_{\{t_j>\lambda\}}}$. This process has a binomial distribution $Bin(n, U_\lambda(t))$. By the empirical central limit theorem (CLT), $B_n(t)$ converges weakly to a $N(0, U_\lambda(t)(1-U_\lambda(t)))$, therefore the finite dimensional distributions are normal for any fixed $t\in[0,1]$. In addition the process is tight due to the uniform continuity of the distribution function U and of $U_\lambda(t)$. This implies that $B_n(t)$ converges weakly to a mean zero gaussian process $Z_{U_\lambda}(t)$. It only remains to find the asymptotic covariance function,

$$Cov(B_n(s), B_n(t)) = Cov[\sqrt{n}(U_{\lambda,n}(s) - U_\lambda(s)), \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))],$$

with $0<s,t<1$. As $U_\lambda(t)$ is constant given $t\in(0,1)$ the covariance function boils down to

$$Cov(B_n(s), B_n(t)) = \frac{n}{(1-U_n(\lambda))^2}Cov(\frac{1}{n}\sum_{i=1}^{n}1_{\{\lambda<t_i\leq s\}}, \frac{1}{n}\sum_{i=1}^{n}1_{\{\lambda<t_i\leq t\}}).$$

The observations $\{t_1,\ldots,t_n\}$ are iid, and therefore $Cov(1_{\{\lambda<t_i\leq s\}}, 1_{\{\lambda<t_j\leq t\}}) = 0$ with $i\neq j$. The covariance function is in this case

$$Cov(B_n(s), B_n(t)) = \frac{1}{(1-U_n(\lambda))^2}Cov(1_{\{\lambda<t_i\leq s\}}, 1_{\{\lambda<t_i\leq t\}}) =$$

$$= \frac{1}{(1-U_n(\lambda))^2}[E(1_{\{\lambda<t_i\leq min(s,t)\}}) - E(1_{\{\lambda<t_i\leq s\}})E(1_{\{\lambda<t_i\leq t\}})] =$$

$$= \frac{(U(min(s,t)) - U(\lambda)) - (U(s)-U(\lambda))(U(t)-U(\lambda))}{(1-U_n(\lambda))^2}, \tag{2.46}$$

with $0<s,t<1$. Therefore $B_n(t)$ converges weakly to the distribution of a mean zero gaussian process $Z_{U_\lambda}(t)$ with covariance function given by

$$Cov(Z_{U_\lambda}(s), Z_{U_\lambda}(t)) = \frac{(min(s,t)-\lambda) - (s-\lambda)(t-\lambda)}{(1-\lambda)^2}.$$

Let F be continuous and strictly increasing and define $u_n = F^{-1}(\lambda)$. Construct $x_1, \ldots, x_n$ *iid* from F via $x_i = F^{-1}(t_i)$ and let $F_n(x)$ denote the empirical distribution function based on $x_1, \ldots, x_n$. By the continuous mapping theorem $\sum_{i=1}^{n} 1_{\{u_n < x_i \le x\}} = \sum_{i=1}^{n} 1_{\{F(u_n) < F(x_i) \le F(x)\}}$ and therefore $F_{u_n,n}(x) = U_{\lambda,n}(t)$. Then $B_n(t)$ can be written as $\sqrt{n}(F_{u_n,n}(y) - F_{u_n}(y))$ with $y = x - u_n$ (see (2.19) and (2.20)) and the covariance function is

$$Cov(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{(F(min(y_1, y_2)) - F(u_n)) - (F(y_1) - F(u_n))(F(y_2) - F(u_n))}{(1 - F(u_n))^2},$$

with $y_1 = F^{-1}(s)$ and $y_2 = F^{-1}(t)$.

Under the null hypothesis $F_{u_n} = GPD_{\xi,\sigma(u_n)}$ the empirical process $S_n(y)$ is written as $\sqrt{n}(F_{u_n,n}(y) - GPD_{\xi,\sigma(u_n)}(y))$ and the covariance function of the limiting process can be written as

$$Cov(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{GPD_{\xi,\sigma}(min(y_1, y_2))}{1 - F(u_n)} - GPD_{\xi,\sigma}(y_1)GPD_{\xi,\sigma}(y_2).$$

Proof of lemma 2.4.1: Let $0 < \alpha < 1$ be the significance level of the test and consider $L(x, F)$ continuous and strictly increasing. By definition

$$P\{T_n > L^{-1}(1 - \alpha, F)\} = \alpha,$$

with $L^{-1}(1 - \alpha, F)$ the $1 - \alpha$ asymptotic quantile.

Consider $L_n(x, Q_n)$ the bootstrap approximation of $L_n(x; F)$ and $L_n^{-1}(1 - \alpha, Q_n)$ its $1-\alpha$ quantile. Therefore if $\sup_{x \in \mathbb{R}} |Q_n(x) - F(x)| \xrightarrow{p} 0$ then $L_n^{-1}(1-\alpha, Q_n) \to L^{-1}(1-\alpha, F)$ with probability one and by Slutsky's theorem

$$P\{T_n > L_n^{-1}(1 - \alpha, Q_n)\} \to P\{T_n > L^{-1}(1 - \alpha, F)\} = \alpha.$$

2.8 Appendix B: Tables

	$t_1(\xi \sim 1)$	$t_5(\xi \sim 0.2)$	$t_{10}(\xi \sim 0.1)$	$t_{30}(\xi \sim 0)$
$\hat{\xi}_n^{Ml}(u_n)$	$[0.70, 1.69]$	$[-0.17, 0.24]$	$[-0.28, 0.39]$	$[-0.43, 0.68]$
$\hat{\xi}_n^{Pi}(u_n^{Pi})$	$[0.29, 1.06]$	$[-0.39, 0.08]$	$[-0.63, -0.06]$	$[-0.64, -0.17]$
$\hat{\xi}_n^{Ml}(x_{(\frac{95}{100}n)})$	$[0.34, 1.75]$	$[0.19, 0.91]$	$[-0.26, 0.33]$	$[-0.28, 0.57]$

Table 2.8.1. *Bootstrap confidence intervals at a significance level* $\alpha = 0.05$ *for different estimators of the tail index:* $\hat{\xi}_n^{Ml}(u_n)$ *with* u_n *estimated by* d^{Wp} *and by* $x_{(\frac{95}{100}n)}$*; and* $\hat{\xi}_n^{Pi}(u_n^{Pi})$ *with* u_n *estimated by* d^{∞}*.* $B = 1000$ *bootstrap samples of size* $n = 1000$ *are drawn from a sequence generated from* t_ν*, with* $\nu = 1,\ 5,\ 10$ *and* 30*.*

	$t_1(\xi \sim 1)$	$t_5(\xi \sim 0.2)$	$t_{10}(\xi \sim 0.1)$	$t_{30}(\xi \sim 0)$
$\hat{\xi}_n^{Ml}(u_n)$	$[0.70, 1.69]$	$[-0.17, 0.24]$	$[-0.28, 0.39]$	$[-0.43, 0.68]$
$\hat{\xi}_n^{Hi}(u_n)$	$[0.82, 1.23]$	$[0.08, 0.37]$	$[-0.42, 0.23]$	$[0.04, 0.20]$

Table 2.8.2. *Bootstrap confidence intervals at a significance level* $\alpha = 0.05$ *for different estimators of the tail index when* u_n *is obtained from the* $GPD_{\xi,\sigma(u_n)}$ *and from the* PD_ξ *respectively. Note* $\hat{\xi}_n^{Ml}(u_n)$ *is* $\hat{\xi}_n^{Hi}(u_n)$ *for the* PD_ξ *case.* $B = 1000$ *bootstrap samples of size* $n = 1000$ *are drawn from a sequence generated from* t_ν*, with* $\nu = 1,\ 5,\ 10$ *and* 30*.*

$n = 1000$	F_1			
F_0	$t_{30}(\xi \sim 0)$	$t_{10}(\xi \sim 0.1)$	$t_5(\xi \sim 0.2)$	$t_1(\xi \sim 1)$
t_{30}	0.06	0.63	0.91	0.97
t_{10}	0.59	0.08	0.79	0.98
t_5	0.95	0.72	0.06	0.99
t_1	0.94	0.94	0.94	0.05

Table 2.8.3. *Empirical power of T_n for a family of t-student distribution functions, with u_n from d^{Wp}. F_0 denotes the data generating process and F_1 the distribution under the alternative hypothesis. Bootstrap replications $B = 1000$, Monte Carlo simulations $m = 500$. Significance level $\alpha = 0.05$.*

F_0	$x_{(700)}$	$x_{(800)}$	$x_{(900)}$	$x_{(950)}$
t_{30}	0.49	0.48	0.43	0.44
t_{10}	0.48	0.48	0.46	0.46
t_5	0.54	0.50	0.48	0.47
t_1	0.64	0.58	0.52	0.48

Table 2.8.4. *Empirical power for a family of t-student distribution functions, with different ad-hoc threshold choices for a sample size $n = 1000$. Bootstrap replications $B = 1000$, Monte Carlo simulations $m = 500$. Significance level $\alpha = 0.05$.*

	$\hat{\xi}_n^{Ml}(u_n)$	$\hat{\xi}_n^{Hi}(u_n)$	$\hat{\xi}_n^{Pi}(u_n^{Pi})$	$\hat{\xi}_n^{Ml}(x_{(\frac{95}{100}n)})$
Dax	$[-0.02; 0.24; 0.84]$	$[0.30; 0.31; 0.36]$	$[-0.50; -0.37; -0.20]$	$[-0.13; 0.22; 0.65]$
Ftse	$[-0.57; -0.26; 0.04]$	$[0.07; 0.11; 0.12]$	$[-0.44; -0.28; -0.08]$	$[-0.54; -0.29; 0.13]$
Ibex	$[-0.12; 0.28; 0.87]$	$[0.32; 0.37; 0.38]$	$[-0.43; -0.21; -0.04]$	$[-0.04; 0.46; 0.90]$
Nikkei	$[-0.13; 0.11; 0.55]$	$[0.33; 0.34; 0.39]$	$[-0.34; -0.19; -0.03]$	$[-0.25; 0.07; 0.50]$
Dow-Jones	$[-0.11; 0.63; 1.52]$	$[0.33; 0.41; 0.44]$	$[-0.24; -0.22; -0.03]$	$[0.05; 0.76; 1.72]$

Table 2.8.5. *Bootstrap confidence intervals ($\alpha = 0.05$) and pointwise estimation of the tail index ξ for stock returns over the period 19/12/1994 – 20/04/2001. Bootstrap samples $B = 1000$.*

VaR	GPD	PD	F_n	Gaussian
Dax	[3.57; 4.16; 7.83]	[3.48; 4.25; 4.93]	[2.96; 4.33; 5.04]	[3.52; 3.62; 3.71]
Ftse	[2.81; 3.04; 3.40]	[2.83; 3.05; 3.31]	[2.83; 3.08; 3.32]	[2.65; 2.78; 2.85]
Ibex	[3.25; 3.92; 4.69]	[2.94; 3.91; 4.62]	[3.02; 4.50; 5.80]	[3.08; 3.19; 3.32]
Nikkei	[3.69; 4.24; 8.30]	[3.33; 4.31; 5.00]	[4.09; 4.73; 5.95]	[3.75; 3.79; 3.83]
Dow-Jones	[1.47; 2.09; 2.60]	[1.56; 2.09; 2.49]	[1.36; 1.90; 2.15]	[1.55; 1.73; 1.97]

Table 2.8.6. *Confidence intervals ($\alpha = 0.05$) and pointwise estimation of the VaR for the different financial returns calculated with different methodologies: our GPD and PD approaches, nonparametric approach F_n, and a parametric approach based on a Gaussian assumption. The VaR indicates the percentage of returns losses with a $p = 0.01$ probability, and a holding period of 1 day. The data covers the period 19/12/1994 – 20/04/2001. Bootstrap samples $B = 1000$.*

Chapter 3

Dependence in the Extremes: A Channel for the transmission of Risk

The extremal index is the key parameter for extending extreme value theory for *iid* random variables to stationary processes, reflecting the level of dependence in the largest observations defined by a threshold sequence $\{u_n\}$. This chapter introduces an estimator for this parameter as the ratio of the number of elements of two point processes defined by a partition of the sample in different blocks, and by the block maxima exceeding the corresponding thresholds $\{v_n\}$ and $\{u_n\}$, with $v_n > u_n$. The estimator is asymptotically unbiased under very general conditions on $\{u_n\}$, consistent (the variance converges to 0 as $n \to \infty$), and the central limit theorem can be applied. Therefore it supports a hypothesis test for the extremal index, and hence for testing the existence of clustering in the extreme values. Other advantages of this method are that it allows some freedom to choose $\{u_n\}$, and it is not very sensitive to the choice of the partition. The analysis of the clustering of the extreme observations in the Frankfurt financial market (DaX Index) sheds some light about the patterns of dependence in financial sequences. The transmission of risk may be due to the dependence found in the largest observations rather than to the dynamics in the volatility process.

3.1 Background

Suppose a random sample from an unknown distribution function F, and let G be the limiting distribution of the sample maximum M_n. Classical Extreme Value Theory shows that under some regularity conditions on the tail of F and for some suitable constants $a_n > 0$, b_n,

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x), \tag{3.1}$$

where G must be of the following types (see de Haan (1976)),

Type I: (Gumbel) $\quad G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty.$

Type II: (Fréchet) $\quad G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\frac{1}{\xi}}} & x > 0, \xi > 0. \end{cases}$

Type III: (Weibull) $\quad G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(-x)^{-\frac{1}{\xi}}} & x < 0, \xi < 0. \end{cases}$

This important result may be extended to study the maximum of a wide class of dependent processes. We concentrate here on stationary sequences where the dependence is restricted by different distributional *mixing* conditions. We distinguish two types of dependence: long range and short range dependence. To limit the first type of dependence we assume a variation of the distributional mixing condition $D(u_n)$ of Leadbetter et al. (1983). Leadbetter's mixing condition is said to hold for a sequence $\{u_n\}$ if for any integers $1 \leq i_1 < \ldots < i_p < j_1 < \ldots < j_{p'} \leq n$ for which $j_1 - i_p \geq l$, we have

$$D(u_n): \quad |F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u_n) - F_{i_1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_{p'}}(u_n)| \leq \alpha_{n,l},$$

where $\alpha_{n,l_n} \rightarrow 0$ as $n \rightarrow \infty$ for some $l_n = o(n)$, and $F_{i_1,\ldots,i_p}(u_n)$ denotes $P\{X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n\}$. Let $D'(u_n)$ be the alternative mixing condition that will be used throughout the paper. This condition is as follows,

$$D'(u_n) : |P\left\{X_{i_1} > u_n \, or \ldots or \, X_{i_p} > u_n \, or \, X_{j_1} > u_n \, or \ldots or \, X_{j_{p'}} > u_n\right\} -$$

$$-P\left\{X_{i_1} > u_n \, or \ldots or \, X_{i_p} > u_n\right\} P\left\{X_{j_1} > u_n \, or \ldots or \, X_{j_{p'}} > u_n\right\} | \leq \alpha_{n,l}. \tag{3.2}$$

Note that these conditions only concern events of the form $\{X_i > u_n\}$ in contrast to more restrictive mixing conditions, for example the strong mixing condition introduced in Rosenblatt (1956). These mixing conditions alone are sufficient to extend the central result given in (3.1) to stationary sequences for some suitable constants not necessarily the ones obtained from the *iid* context. In particular these constants $a_n > 0$, b_n and the extreme value distribution G are the same of the *iid* case under a condition $D''(u_n)$ restricting short range dependence, Leadbetter (1983), that avoids the presence of clusters,

$$D''(u_n) : \quad \underset{n \to \infty}{limsup} \quad n \sum_{j=2}^{[n/k_n]} P\{X_1 > u_n, X_j > u_n\} \to 0 \quad as \quad k_n \to \infty, \tag{3.3}$$

with k_n a sequence that defines a partition of the sample. Otherwise, for a stationary sequence $\{X_n\}$ satisfying only $D'(u_n)$ with $u_n = a_n x + b_n$, we typically have

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \to G^{\theta}(x), \tag{3.4}$$

where θ is the key parameter for extending extreme value theory for *iid* random variables to stationary sequences. This concept, originated in papers by Loynes (1965), O'Brien (1974) and developed in detail by Leadbetter (1983), reflects the effect of the clustering of the observations exceeding u_n on the limiting distribution of the maximum.

There are different interpretations of the extremal index θ, concerning diverse features of the clustering of the largest observations. Loynes (1965) under different mixing conditions found that

$$P\{M_n \leq u_n\} = F^{n\theta}(u_n). \tag{3.5}$$

O'Brien (1987) showed that

$$P\{M_{2,r_n} \le u_n | X_1 > u_n\} \to \theta, \tag{3.6}$$

where M_{2,r_n} is the maximum of $\{X_2, \ldots, X_{r_n}\}$, and $r_n = o(n)$ satisfies certain growth conditions. Note that from this definition of the extremal index it is straightforward to see that $0 \le \theta \le 1$. Alternatively Leadbetter (1983) showed that the inverse of the extremal index is the limiting mean number of exceedances of u_n in an interval of length r_n, *i.e.*

$$E\left[\sum_{j=1}^{r_n} I(X_j > u_n) | \sum_{j=1}^{r_n} I(X_j > u_n) \ge 1\right] \to \theta^{-1}, \tag{3.7}$$

with $I(X > 0)$ the indicator function. By stationarity this is called the limiting mean cluster size of the process. Finally, Hsing (1993) and Ferro and Segers (2003) take advantage of the limiting probability

$$P\{M_n \le u_n\} \to e^{-\theta\tau}, \tag{3.8}$$

with $0 < \tau < \infty$, in two different ways. Hsing approximates the distribution of $n(1-F(M_n))$ by an exponential distribution with mean θ^{-1}, and Ferro and Segers model the process of the interexceedance times defined by u_n by the same limiting exponential distribution.

Expression (3.8) is a transformation of (3.4) where τ is the exponent of an extreme value distribution and $u_n = a_n x + b_n$. In the same way the limiting probability (3.1) may be written as $P\{M_n \le u_n\} \to e^{-\tau}$. Taking logs in this expression, it is immediate to derive that $n(1 - F(u_n)) \to \tau$ for u_n sufficiently high. Then for *iid* sequences, $B_n^{(u_n)} = \sum_{j=1}^{n} I(X_j > u_n)$ converges in distribution to a Poisson random variable with mean τ.

However for dependent stationary sequences where $D''(u_n)$ is not satisfied $B_n^{(u_n)}$ does not converge to a Poisson random variable (the exceedances of u_n are not mutually independent), nevertheless we can define a point process as the result of thinning $B_n^{(u_n)}$. This thinning defines the process $N_{k_n}^{(u_n)}$ formed by the maxima over k_n blocks of length r_n and exceeding u_n, and converges to a Poisson process N with mean $\theta\tau$, see Leadbetter (1983) or Leadbetter et al. (1983). This paper presents an alternative derivation of the extremal index as the result of thinning twice $B_n^{(u_n)}$. The second thinning of $B_n^{(u_n)}$, and hence thinning of $N_{k_n}^{(u_n)}$, defines another point process $N_{k_n}^{(v_n)}$ that converges in distribution to a Poisson process with intensity $\theta^2\tau$. The sequence $\{v_n\}$ satisfies $n(1-F(v_n)) \to \theta\tau$ and is defined by $E[\sum_{j=1}^{r_n} I(X_j > v_n) | \sum_{j=1}^{r_n} I(X_j > u_n) \geq 1] \to 1$. Under some mild conditions on the threshold sequence, this method provides a consistent estimator of the extremal index that outperforms most of the popular estimators and such that it is not very sensitive to the choice of the block size r_n nor the choice of the sequence $\{u_n\}$ in contrast to the rest of the candidates that estimate θ.

The chapter is structured as follows. Section 2 introduces a definition of the extremal index as the ratio of two point processes derived from the asymptotic distribution of the maximum. A natural estimator for this parameter based on these techniques is introduced in Section 3. This section also reviews some of the most popular estimators found in the literature and their statistical properties, in particular bias and variance. The corresponding properties of our estimator are also studied with special emphasis in the analysis of the mean square error of the different methods. The optimal block size selection is also considered and the section concludes with a hypothesis test for the extremal index that is sufficient to test the existence of clustering in the extremes. A simulation experiment for different examples presented in the literature is conducted in Section 4 stressing a Monte-Carlo experiment for the mean square error. Section 5 presents an application to DaX Index returns in order to gain some understanding about the clustering in the extremes and in the volatility of the process. Finally the conclusions are found in Section 6.

3.2 Definition of the extremal index

Suppose throughout that we have n observations from a stationary sequence $\{X_i, i \geq 1\}$ with marginal distribution function F satisfying $[1-F(x)]/[1-F(x^-)] \to 1$ as $x \to \infty$. This condition is sufficient to define a sequence $\{u_n\}$ for each $0 < \tau < \infty$ such that

$$n(1 - F(u_n)) \to \tau. \qquad (3.9)$$

Consider from now on that $\{X_n\}$ satisfies $D'(u_n)$, as defined in (3.2), for each $\tau > 0$. Intuitively this condition gives a measure of the degree of dependence in the process and permits the construction of *almost* independent blocks by the definition of sequences $\{k_n\}$, $\{r_n\}$ with $k_n \to \infty$, $k_n = o(n)$ and $k_n l_n = o(n)$, while r_n is the integer part of n/k_n. The interpretation of these sequences is: k_n is the number of blocks of the sequence of length n, and r_n the size of each block.

Under these assumptions, if $P\{M_n \leq u_n\}$ converges for some $\tau > 0$ then

$$P\{M_n \leq u_n\} \to e^{-\theta\tau}, \tag{3.10}$$

for all $\tau > 0$, with $0 \leq \theta \leq 1$ (see theorem 3.7.1. of Leadbetter et al. (1983) for a detailed proof). The parameter θ is called the extremal index of the sequence $\{X_n\}$ and is the key parameter for extending extreme value theory from *iid* random variables to stationary processes.

Consider $\{k_n\}$, $\{r_n\}$ that define a suitable partition of the sequence $\{X_n\}$, then a sufficient condition for the existence of the extremal index is

$$k_n(1 - F_{1,\ldots,r_n}(u_n)) \to \theta\tau. \tag{3.11}$$

This result is immediate by the approximation of $P\{M_n \leq u_n\}$ by $P^{k_n}\{M_{r_n} \leq u_n\}$ for suitable choices of k_n and r_n, and (3.10) and the linear polynomial expansion of the exponential function. The converse of this result is also true, *i.e.* a stationary sequence with extremal index θ satisfies (3.11) for each $\tau > 0$. The proof is obtained by taking logs in the expression $P^{k_n}\{M_{r_n} \leq u_n\}$ that approximates $e^{-\theta\tau}$.

Consider the number of exceedances of u_n within a block of size r_n. This event defines a sequence of random variables $B_{r_n}^{(u_n)} = \sum_{j=1}^{r_n} I(X_j > u_n)$ for $r_n \to \infty$, and $r_n = o(n)$ whose expected value, by the stationarity of the process, converges to the mean cluster size of the exceedances of u_n in the sequence $\{X_n\}$, that is, the inverse of the extremal index,

$$E\left[B_{r_n}^{(u_n)}|B_{r_n}^{(u_n)} \geq 1\right] \rightarrow \theta^{-1}.$$

This is readily seen since $E\left[B_{r_n}^{(u_n)}|B_{r_n}^{(u_n)} \geq 1\right] = \sum_{j=1}^{\infty} jP\left\{B_{r_n}^{(u_n)} = j|B_{r_n}^{(u_n)} \geq 1\right\} = \frac{r_n P\{X_j > u_n\}}{P\{\bigcup_{j=1}^{r_n}(X_j > u_n)\}}$, and therefore $E\left[B_{r_n}^{(u_n)}|B_{r_n}^{(u_n)} \geq 1\right] = \frac{r_n(1-F(u_n))}{1-F_{1,\ldots,r_n}(u_n)} \rightarrow \theta^{-1}$, if (3.11) holds.

The same argument may be applied to define a process $B_{r_n}^{(v_n)}$ with $v_n \geq u_n$ satisfying

$$E\left[B_{r_n}^{(v_n)}|B_{r_n}^{(u_n)} \geq 1\right] \rightarrow 1. \tag{3.12}$$

It is of interest to note that the sequence $\{v_n\}$ satisfies condition $D'(v_n)$ since $v_n \geq u_n$ and

$$n(1 - F(v_n)) \rightarrow \theta\tau \quad as \quad n \rightarrow \infty. \tag{3.13}$$

In addition, by the structure of dependence (see (3.9) and (3.10)) we have $P\{M_n \leq v_n\} \rightarrow e^{-\theta^2\tau}$. It is immediate now to see that (3.11) holds for the sequence $\{v_n\}$ by

$$k_n(1 - F_{1,\ldots,r_n}(v_n)) \rightarrow \theta^2\tau. \tag{3.14}$$

The event $\{X_i > u_n\}$ and the sequences $\{k_n\}$, $\{r_n\}$ divide the sequence $\{X_n\}$, with extremal index θ, in approximately independent groups of exceedances of u_n where $M_{(j-1)r_n+1,jr_n}$ is the block maxima for $j = 1, \ldots, k_n$. It is clear that the sequence $\{M_{(j-1)r_n+1,jr_n}\}$ is approximately serially independent as n increases if $D'(u_n)$ holds for $\{X_n\}$.

Consider the points j as points in time and define for each n, and k_n, a process $\eta_{k_n}(j/k_n) = M_{(j-1)r_n+1,jr_n}$. The time scale is normalized $t = j/k_n$ on the unit interval $(0, 1]$. Then the exceedances of u_n by the process $\eta_{k_n}(t)$ define a point process $N_{k_n}^{(u_n)}$ on the unit interval (see Kallenberg (1976) for the theory of point processes). Moreover, the point process $N_{k_n}^{(u_n)}$ converges in distribution to a Poisson process N on $(0, 1]$ with intensity parameter $\theta\tau$. To prove this result it is only necessary to show that $E[N_{k_n}^{(u_n)}(a, b]] \rightarrow E[N(a, b]]$ for $0 < a < b \leq 1$ and $P\{N_{k_n}^{(u_n)}(A) = 0\} \rightarrow P\{N(A) = 0\}$ for each finite disjoint union A

of sets $(a_i, b_i] \subset (0,1]$. The proof is analog to the corresponding one found in theorem 4.1. in Leadbetter (1983).

It is interesting to see that the same argument may be applied to construct a thinning of $N_{k_n}^{(u_n)}$ by a sequence $\{v_n\}$ satisfying (3.12). This sequence defines the point process $N_{k_n}^{(v_n)}$ on the unit interval that converges to a Poisson process with intensity measure $\theta^2\tau$. The proof is identical to the case $N_{k_n}^{(u_n)}$ since (3.14) and $D'(v_n)$ hold with $v_n \geq u_n$.

These results provide the setting to define the extremal index as the ratio of the limiting expected value of the point processes $N_{k_n}^{(u_n)}$ and $N_{k_n}^{(v_n)}$,

$$\theta = \lim_{n\to\infty} \frac{E[N_{k_n}^{(v_n)}]}{E[N_{k_n}^{(u_n)}]}. \tag{3.15}$$

The extremal index can also be interpreted as the conditional excess probability of u_n. From the results given in (3.9) and (3.13),

$$\theta = 1 - \lim_{n\to\infty} F_{u_n}(v_n), \tag{3.16}$$

with $F_{u_n}(v_n) = \frac{F(v_n)-F(u_n)}{1-F(u_n)}$. It is clear that as the dependence in the extremes (exceedances of u_n) of the stationary sequence decreases, v_n approaches u_n and θ gets closer to one as for the *iid* case or for weak dependence ($D'(u_n)$ and $D''(u_n)$ hold).

These definitions of the extremal index are also valid for threshold sequences where (3.9) does not hold but the mixing condition in (3.2) still does. Consider $\widetilde{u}_n$ such that $n(1 - F(\widetilde{u}_n)) = \tau_n$, with $\tau_n \to \infty$, and $\tau_n = o(n)$. This condition implies that $P\{M_n \leq \widetilde{u}_n\} \to 0$.

A necessary condition for $\widetilde{u}_n$ in order to define the extremal index in the same way as in (3.15) is that the ratio $\frac{-log P\{M_n \leq \widetilde{u}_n\}}{n(1-F(\widetilde{u}_n))}$ converges to a constant in $(0,1)$. If the sequence $\{X_n\}$ has extremal index θ conditions (3.9) and (3.11) are satisfied for certain sequence u_n. Then, a sufficient condition for $\widetilde{u}_n$ is that

$$\frac{(1 - F(\widetilde{u}_n))(1 - F_{1,\ldots,r_n}(u_n))}{(1 - F(u_n))(1 - F_{1,\ldots,r_n}(\widetilde{u}_n))} \to 1. \tag{3.17}$$

This condition entails this, $k_n(1-F_{1,\ldots,r_n}(\widetilde{u}_n)) = \tau'_n$ with $\tau'_n \to \infty$ and $\tau'_n/\tau_n \to \theta$. The same results that for u_n and τ constant are achieved now for $\widetilde{u}_n$ and τ_n. Therefore, the sequence $B_{r_n}(\widetilde{u}_n)$ satisfies that

$$E\left[B_{r_n}^{(\widetilde{u}_n)} | B_{r_n}^{(\widetilde{u}_n)} \geq 1\right] \to \theta^{-1},$$

and there exists a sequence $\widetilde{v}_n$ such that $n(1-F(\widetilde{v}_n)) = \tau'_n$. Under condition (3.17) for $\{\widetilde{v}_n\}$ instead of $\{\widetilde{u}_n\}$ we obtain that $k_n(1-F_{1,\ldots,r_n}(\widetilde{v}_n)) = \tau''_n$, with $\tau''_n \to \infty$ and $\tau''_n/\tau'_n \to \theta$, and the extremal index may be defined as in (3.15) for the corresponding $\widetilde{u}_n$ and $\widetilde{v}_n$ given that $D'(\widetilde{u}_n)$ holds.

For estimation purposes we will refer to the number of elements of the processes $N_{k_n}^{(\widetilde{u}_n)}$ and $N_{k_n}^{(\widetilde{v}_n)}$ as $Z^*_{\widetilde{u}_n}$ and $Z^*_{\widetilde{v}_n}$ respectively, and $Z_{\widetilde{u}_n}$ and $Z_{\widetilde{v}_n}$ will be used to denote the number of exceedances of $\widetilde{u}_n$ and $\widetilde{v}_n$ by the sequence $\{X_n\}$. Analog notation will be used for the corresponding exceedances of u_n and v_n. Note the variables $Z^*_{\widetilde{u}_n}$ and $Z^*_{u_n}$ can be interpreted as the number of blocks of the partition defined by $\{k_n\}$, $\{r_n\}$ where there is at least one exceedance of $\widetilde{u}_n$ and u_n respectively.

3.3 Estimation of the extremal index

The extremal index represents the clustering of the largest observations determined by a sequence $\{u_n\}$ sufficiently high to satisfy a condition of type (3.9). The serial dependence in these observations has an effect on the distribution of the maximum of the stationary sequence, that is, $P\{M_n \leq u_n\}$ is $F^{n\theta}(u_n)$ instead of $F^n(u_n)$ for n and u_n sufficiently large.

This result leads to the first estimator of the extremal index for appropriate sequences k_n, r_n satisfying that $P^{k_n}\{M_{r_n} \leq u_n\}$ approximates $P\{M_n \leq u_n\}$. Then, by taking logs in both expressions, $\theta = \frac{logP\{M_{r_n} \leq u_n\}}{r_n logF(u_n)}$. A natural estimator for the extremal index is in this case,

$$\hat{\theta}_n^{(1)} = \frac{log(1-Z^*_{u_n}/k_n)}{r_n log(1-Z_{u_n}/n)}, \tag{3.18}$$

with the notation introduced in the last section. The ratio Z_{u_n}/n is an estimator of $1-F(u_n)$, and $Z^*_{u_n}/k_n$ an estimator of $1-F_{1,\ldots,r_n}(u_n)$.

On the other hand the concept of extremal index introduced by Leadbetter (1983), θ^{-1} the limiting mean cluster size of the exceedances, yields this estimator

$$\hat{\theta}_n^{(2)} = \frac{Z^*_{u_n}}{Z_{u_n}}. \tag{3.19}$$

This method is called the blocks method and may be considered a simplified version of $\hat{\theta}_n^{(1)}$. Another popular method is the runs estimator, that may be seen as the estimator of the extremal index for the definitions introduced in O'Brien (1987) or in Hsing (1993),

$$\overline{\theta}_n = \frac{W_{u_n}}{Z_{u_n}}, \tag{3.20}$$

where $W_{u_n} = \sum_{i=1}^{n-r_n} I(X_i > u_n)(1 - I(X_{i+1} > u_n)) \cdot \cdot (1 - I(X_{i+r_n} > u_n))$.

Our definition of the extremal index yields an appealing estimator of θ given by the ratio of $Z^*_{v_n}$ and $Z^*_{u_n}$ or alternatively $Z^*_{\tilde{v}_n}$ and $Z^*_{\tilde{u}_n}$. For u_n and v_n sequences satisfying (3.9) and (3.13) our estimator $\widetilde{\theta}_n$ is given by

$$\widetilde{\theta}_n = \frac{Z^*_{v_n}}{Z^*_{u_n}}, \tag{3.21}$$

representing the corresponding thinnings defined by the sequence k_n and the thresholds u_n and v_n. The estimator, however, is not fully specified since these sequences are not determined. By (3.9) an appropriate candidate for this threshold sequence is given by extreme order statistics (see section 2.5. in Leadbetter et al. (1983)). In turn an adequate choice of v_n is given by the order statistic of the stationary sequence $\{X_n\}$ satisfying the empirical counterpart of (3.12), *i.e.*

$$v_n = \max_{1\le i\le n} \left\{ x_i, i = 1, \ldots, n \middle| \frac{1}{Z^*_{u_n}} \sum_{j=1}^{k_n} B^{(x_i)}_{r_n,j} = 1 \right\}, \tag{3.22}$$

with $B^{(u_n)}_{r_n,j} = \sum_{k=(j-1)r_n+1}^{jr_n} I(X_k > u_n)$. This expression boils down to $v_n = x_{(n-Z^*_{\tilde{u}_n})}$, extreme order statistic, with $x_{(1)} \leq \ldots \leq x_{(n)}$ the sequence of order statistics. By (3.17) the corresponding expressions apply to $\widetilde{u}_n$ and $\widetilde{v}_n$ being intermediate order statistics.

If the threshold u_n is estimated by an extreme order statistic the point process $N^{(u_n)}_{k_n}$ converges to a Poisson process, and its variance in consequence converges to a constant. This is a serious inconvenient for the consistency of the majority of the estimators of θ that is overcome in our setup by using $\widetilde{u}_n$ (intermediate order statistic).

3.3.1 Statistical properties of the different estimators

Consider first the case of $\widetilde{\theta}_n$ as the quotient of the random variables $Z^*_{v_n}$ and $Z^*_{u_n}$ where v_n and u_n satisfy (3.13) and (3.9) respectively, that is,

$$\widetilde{\theta}_n = \frac{Z^*_{v_n}}{Z^*_{u_n}}.$$

By the second order Taylor expansion of $E[Z^*_{v_n}/Z^*_{u_n}]$ about the respective expected values (delta method) we have that

$$E[\widetilde{\theta}_n] = \frac{E[Z^*_{v_n}]}{E[Z^*_{u_n}]}\left(1 + \frac{V[Z^*_{u_n}]}{E[Z^*_{u_n}]^2} - \frac{Cov[Z^*_{v_n}, Z^*_{u_n}]}{E[Z^*_{u_n}]E[Z^*_{v_n}]}\right) + O(\frac{1}{\tau^2}). \tag{3.23}$$

The different contributions to $Z^*_{u_n}$ are not mutually independent. In particular, $E[Z^{*2}_{u_n}] = k_n P\{M_1 > u_n\} + \sum_{i=1}^{k_n}\sum_{i\neq j}^{k_n} P\{M_i > u_n, M_j > u_n\}$, where M_i is used to denote the maximum of $\{X_{(i-1)r_n+1}, \ldots, X_{ir_n}\}$. By stationarity the variance can be expressed as $V[Z^*_{u_n}] = E[Z^*_{u_n}] + k_n^2 P\{M_1 > u_n, M_2 > u_n\} - E^2[Z^*_{u_n}] - k_n P\{M_1 > u_n, M_2 > u_n\}$. Under $D'(u_n)$ the difference between $k_n^2 P\{M_1 > u_n, M_2 > u_n\}$ and $E^2[Z^*_{u_n}]$ converges to 0 as n increases, and $k_n P\{M_1 > u_n, M_2 > u_n\}$ is well approximated by $E[Z^*_{u_n}]P\{M_1 > u_n\}$ that in turn also converges to 0. The covariance takes a similar expression, $Cov[Z^*_{u_n}, Z^*_{v_n}] = E[Z^*_{v_n}] + k_n^2 P\{M_1 > u_n, M_2 > v_n\} - E[Z^*_{u_n}]E[Z^*_{v_n}] - k_n P\{M_1 > u_n, M_2 > v_n\}$ that boils down to $Cov[Z^*_{u_n}, Z^*_{v_n}] = E[Z^*_{v_n}]$.

Therefore expression (3.23) for n sufficiently high is as follows

$$E[\widetilde{\theta}_n] = \frac{E[Z^*_{v_n}]}{E[Z^*_{u_n}]}\left(1 + \frac{E[Z^*_{u_n}]}{E[Z^*_{u_n}]^2} - \frac{E[Z^*_{v_n}]}{E[Z^*_{u_n}]E[Z^*_{v_n}]}\right) + O(\frac{1}{\tau^2}),$$

and it is immediate to see that the expected value of our estimator takes this expression,

$$E[\widetilde{\theta}_n] = \theta + O(\frac{1}{\tau^2}). \tag{3.24}$$

For the analysis of the variance it is useful to derive the conditional moments. Consider $Z^*_{u_n} = z^*_{u_n}$ known, and note that the sequences u_n and v_n are related by this expression,

$$1 - F_{1,\ldots,r_n}(v_n) = (1 - F_{1,\ldots,r_n}(u_n))\left(1 - \frac{F_{1,\ldots,r_n}(v_n) - F_{1,\ldots,r_n}(u_n)}{1 - F_{1,\ldots,r_n}(u_n)}\right). \tag{3.25}$$

Then by (3.11), $E[\widetilde{\theta}_n|Z^*_{u_n} = z^*_{u_n}] = 1 - \frac{F_{1,\ldots,r_n}(v_n) - F_{1,\ldots,r_n}(u_n)}{1-F_{1,\ldots,r_n}(u_n)}$, and the conditional variance takes this form

$$V[\widetilde{\theta}_n|Z^*_{u_n} = z^*_{u_n}] = \frac{1}{z^*_{u_n}}\left(1 - \frac{F_{1,\ldots,r_n}(v_n) - F_{1,\ldots,r_n}(u_n)}{1 - F_{1,\ldots,r_n}(u_n)}\right). \tag{3.26}$$

By the law of iterated expectations the unconditional variance can be decomposed in two different terms, $V[\widetilde{\theta}_n] = V[E[\widetilde{\theta}_n|Z^*_{u_n}]] + E[V[\widetilde{\theta}_n|Z^*_{u_n}]]$. It is clear the first term is 0, and by the Taylor expansion of $E[1/Z^*_{u_n}]$ about $E[Z^*_{u_n}]$ we obtain that

$$E[V[\widetilde{\theta}_n|Z^*_{u_n} = z^*_{u_n}]] = \left(1 - \frac{F_{1,\ldots,r_n}(v_n) - F_{1,\ldots,r_n}(u_n)}{1 - F_{1,\ldots,r_n}(u_n)}\right)\left(\frac{1}{E[Z^*_{u_n}]} + \frac{V[Z^*_{u_n}]}{E^3[Z^*_{u_n}]}\right). \tag{3.27}$$

In consequence,

$$V[\widetilde{\theta}_n] = \left(1 - \frac{F_{1,\ldots,r_n}(v_n) - F_{1,\ldots,r_n}(u_n)}{1 - F_{1,\ldots,r_n}(u_n)}\right)\left(\frac{1}{\theta\tau} + O\left(\frac{1}{\tau^2}\right)\right) = O\left(\frac{1}{\tau}\right). \tag{3.28}$$

Therefore the mean square error (MSE) of our estimator is of order $O(\frac{1}{\tau})$ with τ constant. This result implies that this estimator is not consistent for u_n and v_n defined by extreme order statistics. The consistency, however, will be achieved when these sequences are replaced by $\widetilde{u}_n$ and $\widetilde{v}_n$ intermediate order statistics as it is shown in the following section.

Our estimator may be interpreted as a refinement of the standard blocks method $\hat{\theta}_n^{(2)}$ by writing $\widetilde{\theta}_n = \frac{Z^*_{v_n}/Z_{u_n}}{\hat{\theta}_n^{(2)}}$. The asymptotic properties of the latter estimator $\hat{\theta}_n^{(2)}$ are derived in Hsing (1991) or in Smith and Weissman (1994). By means of the delta method they find that $E[\hat{\theta}_n^{(2)}] = \theta + O(\frac{1}{\tau})$, and the variance is $V[\hat{\theta}_n^{(2)}] = O(\frac{1}{\tau})$. Therefore the bias of this estimator is higher than the bias of $\widetilde{\theta}_n$, but the mean square error (MSE) of both estimators is $O(1/\tau)$.

For the logs method,

$$E[\hat{\theta}_n^{(1)}] = \frac{E[Z^*_{u_n}]}{E[Z_{u_n}]}\left(1 + \frac{E[Z^*_{u_n}]}{2k_n} + \frac{E^2[Z^*_{u_n}]}{6k_n^2}\right) = \theta + O\left(\frac{\tau}{k_n}\right), \ \ and \ \ V[\hat{\theta}_n^{(1)}] = O(\frac{1}{\tau}).$$

This estimator is asymptotically unbiased, but it is not consistent either for τ constant.

3.3.2 Inference for the Extremal Index

Consider now the sequences $\widetilde{u}_n$ and $\widetilde{v}_n$ defined by the conditions $\tau'_n = k_n(1-F_{1,\ldots,r_n}(\widetilde{u}_n))$, $\tau''_n = k_n(1 - F_{1,\ldots,r_n}(\widetilde{v}_n))$, with $\tau'_n \to \infty$, $\tau''_n \to \infty$ and $\tau''_n/\tau'_n \to \theta$. In this case the first two moments of the random variables $Z^*_{\widetilde{u}_n}$ and $Z^*_{\widetilde{v}_n}$ diverge to infinity. By stationarity the variance is given by this expression,

$$V[Z^*_{\widetilde{u}_n}] = E[Z^*_{\widetilde{u}_n}] + \left(k_n^2 P\{M_1 > \widetilde{u}_n, M_2 > \widetilde{u}_n\} - E^2[Z^*_{\widetilde{u}_n}]\right) - k_n P\{M_1 > \widetilde{u}_n, M_2 > \widetilde{u}_n\}.$$

Note that in this case, under $D'(u_n)$ for n sufficiently high, the variance is

$$V[Z^*_{\widetilde{u}_n}] = E[Z^*_{\widetilde{u}_n}] - E[Z^*_{\widetilde{u}_n}]P\{M_1 > \widetilde{u}_n\}. \tag{3.29}$$

The covariance in turn takes this expression,

$$Cov[Z^*_{\widetilde{u}_n}, Z^*_{\widetilde{v}_n}] = E[Z^*_{\widetilde{v}_n}] - E[Z^*_{\widetilde{v}_n}]P\{M_1 > \widetilde{u}_n\}.$$

Therefore expression (3.23) is as follows

$$E[\widetilde{\theta}_n] = \frac{\tau_n''}{\tau_n'}\left(1 + \frac{\tau_n' P\{M_1 \leq \widetilde{u}_n\}}{(\tau_n')^2} - \frac{\tau_n'' P\{M_1 \leq \widetilde{u}_n\}}{\tau_n' \tau_n''}\right) + o\left(\frac{1}{\tau_n}\right), \tag{3.30}$$

that boils down to $E[\widetilde{\theta}_n] = \theta + o\left(\frac{1}{\tau_n}\right)$, by the definition of τ_n' and τ_n''.

This estimator of θ is now asymptotically unbiased, and for $\tau_n < k_n$, $\tau_n^2 > k_n$, $\widetilde{\theta}_n$ outperforms $\hat{\theta}_n^{(1)}$ in this sense. For $\tau_n > k_n$ this result is trivial.

In order to find the unconditional variance in this case, we calculate first the conditional moment.

$$V[\widetilde{\theta}_n | Z^*_{\widetilde{u}_n} = z^*_{\widetilde{u}_n}] = \frac{1}{z^{*2}_{\widetilde{u}_n}} V[Z^*_{\widetilde{v}_n} | Z^*_{\widetilde{u}_n} = z^*_{\widetilde{u}_n}]. \tag{3.31}$$

Applying (3.29) to the random variable $Z^*_{\widetilde{v}_n} | Z^*_{\widetilde{u}_n}$,

$$V[Z^*_{\widetilde{v}_n} | Z^*_{\widetilde{u}_n} = z^*_{\widetilde{u}_n}] = E[Z^*_{\widetilde{v}_n} | Z^*_{\widetilde{u}_n} = z^*_{\widetilde{u}_n}] P\{M_1 \leq \widetilde{v}_n | M_1 > \widetilde{u}_n\},$$

that amounts to

$$V[Z^*_{\widetilde{v}_n} | Z^*_{\widetilde{u}_n} = z^*_{\widetilde{u}_n}] = z^*_{\widetilde{u}_n} P\{M_1 > \widetilde{v}_n | M_1 > \widetilde{u}_n\} P\{M_1 \leq \widetilde{v}_n | M_1 > \widetilde{u}_n\}. \tag{3.32}$$

Then, in the same way as in (3.27),

$$V[\widetilde{\theta}_n] = \left(1 - \frac{F_{1,\ldots,r_n}(\widetilde{v}_n) - F_{1,\ldots,r_n}(\widetilde{u}_n)}{1 - F_{1,\ldots,r_n}(\widetilde{u}_n)}\right)\left(\frac{F_{1,\ldots,r_n}(\widetilde{v}_n) - F_{1,\ldots,r_n}(\widetilde{u}_n)}{1 - F_{1,\ldots,r_n}(\widetilde{u}_n)}\right)\left(\frac{1}{E[Z^*_{\widetilde{u}_n}]} + \frac{V[Z^*_{\widetilde{u}_n}]}{E^3[Z^*_{\widetilde{u}_n}]}\right),$$

that in turn is

$$V[\widetilde{\theta}_n] = \left(1 - \frac{F_{1,\ldots,r_n}(\widetilde{v}_n) - F_{1,\ldots,r_n}(\widetilde{u}_n)}{1 - F_{1,\ldots,r_n}(\widetilde{u}_n)}\right)\left(\frac{F_{1,\ldots,r_n}(\widetilde{v}_n) - F_{1,\ldots,r_n}(\widetilde{u}_n)}{1 - F_{1,\ldots,r_n}(\widetilde{u}_n)}\right)\frac{1}{\tau_n'} + o\left(\frac{1}{\tau_n}\right). \tag{3.33}$$

Under $D'(\widetilde{u}_n)$ the distribution of $Z^*_{\widetilde{v}_n} | Z^*_{\widetilde{u}_n}$ is well approximated ($\sim$) by a binomial distribution with parameters $bin\left(Z^*_{\widetilde{u}_n}, 1 - \frac{F_{1,\ldots,r_n}(\widetilde{v}_n) - F_{1,\ldots,r_n}(\widetilde{u}_n)}{1 - F_{1,\ldots,r_n}(\widetilde{u}_n)}\right)$, and $Z^*_{\widetilde{u}_n}$ by a $bin\left(k_n, 1 - F_{1,\ldots,r_n}(\widetilde{u}_n)\right)$. Then, the distribution of $\widetilde{\theta}_n$ can be approximated by a normal distribution with parameters given in (3.30) and (3.33).

On the other hand the relation between the tails introduced in (3.25) holds for $\widetilde{u}_n$ and $\widetilde{v}_n$, and by assumption $\tau''_n/\tau'_n \to \theta$, yielding that $1 - \frac{F_{1,\ldots,r_n}(\widetilde{v}_n) - F_{1,\ldots,r_n}(\widetilde{u}_n)}{1-F_{1,\ldots,r_n}(\widetilde{u}_n)} \to \theta$. In turn, the distribution of $\widetilde{\theta}_n$ is approximated by

$$\widetilde{\theta}_n \overset{w}{\sim} N\left(\theta, \frac{\theta(1-\theta)}{\tau'_n}\right). \tag{3.34}$$

By the structure of dependence $\frac{\tau'_n}{\tau_n} \to \theta$ as n goes to infinity, and hence $\widetilde{\theta}_n \overset{w}{\sim} N\left(\theta, \frac{1-\theta}{\tau_n}\right)$ results a valid approximation for the distribution of $\widetilde{\theta}_n$. More formally, we can obtain a test statistic that is asymptotically parameter free,

$$T_n = \frac{\widetilde{\theta}_n - \theta}{\sqrt{1-\theta}}\sqrt{\tau_n} \overset{w}{\to} N(0,1). \tag{3.35}$$

The asymptotic confidence intervals for θ are easily calculated from the former expression.

$$\theta \in \left[\widetilde{\theta}_n \pm z_{1-\alpha/2}\sqrt{\frac{1-\widetilde{\theta}_n}{\tau_n}}\ \right], \tag{3.36}$$

with $z_{1-\alpha/2}$ the quantile of the standard normal distribution. This interval is an approximation of the true confidence interval for finite samples. The exact confidence region for small sample sizes may be better approximated by resampling techniques. The confidence interval takes this expression

$$\theta \in \left[\widetilde{\theta}_n - \sqrt{\frac{1-\widetilde{\theta}_n}{\tau_n}} J_n^{-1}(1-\frac{\alpha}{2}, F), \widetilde{\theta}_n - \sqrt{\frac{1-\widetilde{\theta}_n}{\tau_n}} J_n^{-1}(\frac{\alpha}{2}, F)\right], \tag{3.37}$$

where $J_n^{-1}(1-\alpha, F)$ is the $1-\alpha$ quantile of the sampling distribution $J_n(F)$ of the statistic T_n. In practice this quantile is approximated by the order statistic $T_{n,((1-\alpha)B)}$ of the sample $T_{n,1}, \ldots, T_{n,B}$ with B the number of iterations. The notation F in the distribution $J_n(F)$ refers to Monte Carlo simulation, that is, the generating process of the data is known. Otherwise $J_n(F)$ must be approximated by $J_b(F)$ with $b < n$, $b/n \to 0$ (subsampling), or by $J_n(F^*)$ with F^* representing blocks bootstrap methods. The naïve bootstrap does not work in this context due to the serial dependence in the data. Nevertheless, as it is seen in the simulations, the gaussian asymptotic intervals give reliable approximations of the exact confidence regions for moderate sample sizes.

Our interest however lies on testing hypotheses of the type $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$. This one-sided hypothesis test permits to assess the mixing condition $D''(u_n)$ introduced in (3.3) by imposing $\theta_0 = 1$. In other words, if $\theta = 1$ there are no clusters of extreme values (exceedances of some threshold) in the stationary sequence. The null hypothesis amounts to see if θ_0 is contained in the interval $\left(-\infty, \widetilde{\theta}_n + z_{1-\alpha}\sqrt{\frac{1-\widetilde{\theta}_n}{\tau_n}}\right]$ or alternatively in $\left(-\infty, \widetilde{\theta}_n - \sqrt{\frac{1-\widetilde{\theta}_n}{\tau_n}} J_n^{-1}(\alpha, F)\right]$, the bootstrap approximation.

Consider the example due to Chernick (1981) for $\{X_n\}$ a strictly stationary first order autoregressive sequence driven by the model $X_i = \frac{1}{r}X_{i-1} + \varepsilon_i$, where $r \geq 2$ is an integer, ε_i are discrete uniforms on $\{0, 1/r, \ldots, (r-1)/r\}$, being independent of X_{i-1}, and X_i having a uniform distribution on $[0, 1]$. The extremal index is given by $\theta = \frac{r-1}{r}$. The plot given in figure 3.7.1 describes the curve of the estimates of θ for different partitions and different sample sizes. The upper panel is for $n = 200$ and the lower panel considers $n = 1000$.

Two conclusions stem from these plots. First, it is clear that condition $D''(\hat{u}_n)$ is rejected with $\alpha = 0.05$, and second, the confidence intervals for θ are smaller as n increases. This is caused by the choice of an increasing order statistic, $\hat{u}_n = x_{(n-k)}$ with $k = \sqrt{2n}$ as threshold.

3.3.3 Some comments on the block size selection

The partition of the sequence $\{X_n\}$ in k_n blocks of size r_n has two main features: First, it defines a point process $N_{k_n}^{(u_n)}$ that converges to a serially independent process, and second, the distribution of this sequence converges to a Poisson process as k_n goes to infinity. The majority of the estimators for the extremal index found in the literature are tied to that partition. This dependence turns explicit for example for the logs method where k_n appears in the expression for $\hat{\theta}_n^{(1)}$, as well as in its expected value.

If the observations are independent or weak dependent, $N_{k_n}^{(u_n)}$ with $k_n = n$ defines itself an *iid* point process and $\theta = 1$. Otherwise the extremal index is less than 1 and the partition $k_n < n$ plays a central role in the estimation of the extremal index.

Provided n, an adequate choice of the sequence k_n along with a suitable threshold u_n define a sequence given by the maxima over the corresponding blocks of observations. Testing condition $D'(u_n)$ in practice is replaced by testing for serial independence in this sequence of maxima. Hypothesis tests for the latter condition require large sample sizes and most of them rely on gaussian assumptions. A naïve alternative to these methods is dropping the first partitions defined by $k_n = n, n-1, \ldots$, that have a high likelihood of entailing serial dependence in $N_{k_n}^{(u_n)}$, and analyzing the performance of the estimators by the stability of the corresponding estimates along the different partitions.

The influence of the choice of k_n on the estimates of the extremal index also depends on the choice of the threshold u_n. For example in the blocks method the estimates of θ are driven by the corresponding partition for low u_n. In this case $Z_{u_n}^*$ and k_n take the same values resulting in a sequence of estimates that approaches 0 when k_n decreases.

A similar situation occurs for the logs method with u_n a low threshold. The numerator in this case collapses to $-\infty$ and the estimator is not defined for the corresponding partition. These effects are not present in $\widetilde{\theta}_n$ because a larger ratio $Z_{u_n}^*/k_n$ given by a low threshold is compensated by a large value of $Z_{v_n}^*/k_n$, that is, the sequence v_n varies according to u_n.

3.4 Simulations: Some examples

We now consider some examples from the literature showing short range dependence $(0 < \theta < 1)$ reflected by a distribution of the maximum satisfying (3.5).

The following example is the doubly stochastic model studied by Smith and Weissman (1994). Let $\{\xi_i, i \geq 1\}$ be *iid* with distribution function F, and suppose that $Y_1 = \xi_1$, and for $i > 1$, $Y_i = Y_{i-1}$ with probability ψ, and $Y_i = \xi_i$ with probability $1 - \psi$. The doubly stochastic sequence $\{X_i\}$ is defined by $X_i = Y_i$ with probability η, and $X_i = 0$ with probability $1-\eta$, with these different events mutually independent. The extremal index is

$$\theta = \frac{1-\psi}{1-\psi+\psi\eta}.$$

The following pictures, figure (3.7.2), represent the paths of the different estimators for the extremal index. The threshold sequence is estimated by an order statistic: $\hat{u}_n = x_{(n-k)}$. We have implemented two different types of order statistics for samples of size $n = 200$ and $n = 1000$ observations. An extreme order statistic ($k = 20$ fixed), and an intermediate order statistic ($k = \sqrt{2n}$). We present only the estimates of θ for the intermediate order statistic since the other threshold sequence provides similar results for these sample sizes in this example.

The curves describe the sample means of the different estimates of the extremal index and for different partitions of the sample for $m = 100$ simulated sequences generated from the model introduced in Smith and Weissman with $\psi = 0.9$ and $\eta = 0.7$. Suppose also F, a Fréchet distribution $F(x) = \exp(-x^{-\alpha})$ with $\alpha = 1$ and $x \in (0, \infty)$.

The confidence intervals for $\widetilde{\theta}_n$ derived in the last section are not plotted. Instead we have represented the simulated standard deviation of the different estimators for $m = 100$ in order to present a fair comparison between the three competitors. The standard deviation for the different partitions is estimated via Monte Carlo simulation by $\hat{\sigma}_{k_n}$ with

$$\hat{\sigma}^2_{k_n} = \frac{1}{m-1}\sum_{i=1}^{m}(\theta_{i,est} - \overline{\theta}_{est})^2, \tag{3.38}$$

and $\overline{\theta}_{est}$ the sample mean of the different estimates.

Apparently the blocks method is the best method. After the first partitions of the sample the curve of the estimates of θ remains stable very close to the target line for the three methods. Nevertheless, the blocks method estimator has smaller variance. In addition, focusing on figure 3.7.3 it is clear that the different estimators of θ analyzed in this example are consistent and the blocks method is more efficient. The mean square error is estimated from the simulated sequences generated for figure 3.7.2, and takes this expression,

$$MSE(\theta_{est}) = \frac{1}{m}\sum_{i=1}^{m}(\theta_{i,est} - \theta)^2.$$

These results agree with the conclusions found in Smith and Weissman (1994). However the impressive performance of the blocks method may be due to the low value of the extremal index ($\theta = 0.137$) and the choice of a low threshold estimate. Under these circumstances, the curve of estimates of θ by the blocks method is decaying as k_n decreases (r_n increases) and approaches the true parameter. To get an insight into this, we also study a doubly stochastic process where the extremal index is significantly higher. Suppose $\psi = 0.5$ and $\eta = 0.5$, *i.e.* $\theta = 0.66$. The plots displayed in figures 3.7.4 and 3.7.5 are the analogs of figures 3.7.2 and 3.7.3.

The blocks method in this example does not work. The number of blocks with an exceedance of $\hat{u}_n$ ($Z^*_{\hat{u}_n}$) is similar to k_n for each partition. Therefore the estimator decreases as k_n decreases since $Z_{\hat{u}_n}$ remains constant. On the other hand the logs method improves as n increases and the mean square error of $\widetilde{\theta}_n$ and $\hat{\theta}_n^{(1)}$ are negligible for $n = 1000$.

Finally the exact and asymptotic confidence intervals for θ are displayed in figure 3.7.6 to assess the estimates given by $\widetilde{\theta}_n$.

3.5 Clustering in Financial Series: The Case of DaX Index

Financial returns are characterized by a series of stylized facts: leverage effect (after periods of high volatility the likelihood of losses is higher than in calm periods), heavy tails, clustering of the largest observations and some skewness towards the losses tail. The seminal paper of Engle and Bollerslev (1986) proposed the popular GARCH models, Generalized Auto-Regressive Conditional Heteroscedastic volatility models, to explain these features of the data. In general, the GARCH(1,1) is sufficient to model most of the financial returns.

It takes this expression,

$$X_i = \varepsilon_i \sigma_i, \qquad \sigma_i^2 = \omega + \alpha X_{i-1}^2 + \beta \sigma_{i-1}^2,$$

with ω, α, $\beta > 0$, and $\alpha + \beta < 1$, that can be interpreted as an ARMA(1,1) model for the squares,

$$X_i^2 = \omega + (\alpha + \beta) X_{i-1}^2 + \nu_i - \beta \nu_{i-1},$$

with $\nu_i = \sigma_i^2 (\varepsilon_i^2 - 1)$.

According to this model, the dependence found in the financial returns is driven by the second moments. The literature concerning this topic is enormous; up to the extent that there exist different GARCH type models to explain particular characteristics of the financial series.

We propose to analyze some of these stylized facts, in particular the clustering of the largest observations, by means of the extremal index. A value of θ significatively less than 1 shows certain short range dependence reflected in the clustering of the largest observations. This may be interpreted as a pattern in the occurrence of the extreme values, that is, once a large loss in the asset return has occurred we can expect a period of large losses (values exceeding some threshold). The average length of this period is the inverse of the extremal index.

The data we use to illustrate this methodology consists on the analysis of the Frankfurt financial market (DaX Index) over the period 19/12/1994 – 20/04/2001. These data have been collected from *www.freelunch.com*. The observations considered for the analysis are the logarithmic returns measured in percentage terms and denoted as r_t,

$$r_t = 100\,(log\ P_t - log\ P_{t-1}),$$

with P_t the original prices at time t. In figure 3.7.7 DaX Index returns, r_t, and the corresponding sequence of squared returns, r_t^2, are plotted.

The analysis of the extremal index for both tails (figure 3.7.8) shows certain clustering in the occurrence of the positive and negative extreme values. The confidence intervals derived from $\widetilde{\theta}_n$ do not contain $\theta = 1$ for $\alpha = 0.05$.

These pictures also depict a higher level of clustering for the largest negative returns than for the positive values. This fact can be statistically tested by means of a confidence interval for the difference of the extremal indexes corresponding to the positive and negative tail. This confidence interval takes this expression

$$\theta_{pos} - \theta_{neg} \in \left[\widetilde{\theta}_{n,pos} - \widetilde{\theta}_{n,neg} \pm z_{1-\alpha/2} \sqrt{\frac{1-\widetilde{\theta}_{n,pos}}{\tau_{n,pos}} + \frac{1-\widetilde{\theta}_{n,neg}}{\tau_{n,neg}}} \right]. \tag{3.39}$$

It is important to mention that $\widetilde{\theta}_{n,pos}$ and $\widetilde{\theta}_{n,neg}$ are considered independent. This can lead to obtain smaller confidence intervals given α compared to considering dependent estimators with positive correlation. For some partitions of the sample it is statistically significant that the clustering for the positive extreme values is smaller than for the largest negative returns (figure 3.7.9).

The analysis of the clustering of the largest values for the sequence of the volatility (squres of returns) deserves some interesting comments. The confidence interval introduced in (3.39) may be applied to test the difference between the extremal index of the volatility sequence θ_{vol} and θ_{pos} or θ_{neg} (figure 3.7.10). The results derived from both tests, $\theta_{pos} - \theta_{vol}$ and $\theta_{neg} - \theta_{vol}$, point out that the extreme values of the volatility sequence are driven by the negative extreme values. Therefore these observations are bigger in absolute value than the largest positive returns. This fact explains the negative skewness of the returns sequence.

Finally it is worth mentioning the stylized fact of heavy tails. By Berman's condition (Berman, 1964), if $\{r_t\}$ is a standard normal sequence and $Cov(r_t, r_{t-j})\ log\ j \to 0$ as $j \to \infty$, the extremal index of the sequence is $\theta = 1$. In practice, the autocorrelation function of the returns of a financial series is usually close to zero, also in this case and then the second part of Berman's condition holds. Therefore, if $\theta < 1$ the sequence of the returns of the DaX Index is not normally distributed but heavy tailed. This suggests that the existence of clustering of the extreme values in a financial series implies that the distribution of the observations is heavy tailed. Hence it is not sufficient with the second moments of $\{r_t\}$ to know the structure of dependence of the sequence. Moreover, the dependence in the extremes plays an important role and this dependence stems from the heavy tails.

3.6 Conclusion

The aim of this chapter has been to propose an estimator for the extremal index defined by the ratio of the number of exceedances of two threshold sequences. This estimator possesses two appealing properties: First, it is not necessary to choose a sequence $\{u_n\}$ satisfying the Poisson condition in the limit, and second it is not very sensitive to the block size selection.

Regarding the asymptotic properties of our estimator, we can conclude that are estimator has the same order of convergence than the standard methods (the respective variances are of the same order). However, under very general conditions our estimator is asymptotically unbiased outperforming the other two methods that are not free from a residual term. Our estimator also works better than these methods in two manners: it is not so dependent of the corresponding partition of the sequence, and it relaxes the selection of the threshold sequence.

In addition, the absence of dependence on the Poisson condition permits to propose a hypothesis test for the extremal index. We find this test useful in different ways: it formally assesses the estimates of the extremal index, it introduces an innovative procedure for testing the existence of clustering in the occurrence of extreme events, and it may be useful to determine the skewness and kurtosis of the distribution of the data by testing the difference of extremal indexes between both tails.

Finally, the application of these methodologies to financial series (DaX Index) confirms the existence of short range dependence in the extreme observations; that is, some clustering of the extreme values of the positive and negative returns. The clustering is higher for the negative tail. By Berman's condition, the distribution of the observations is heavy tailed since θ is statistically less than 1. These results agree with the stylized facts found in most of the financial series.

3.7 Appendix: Figures

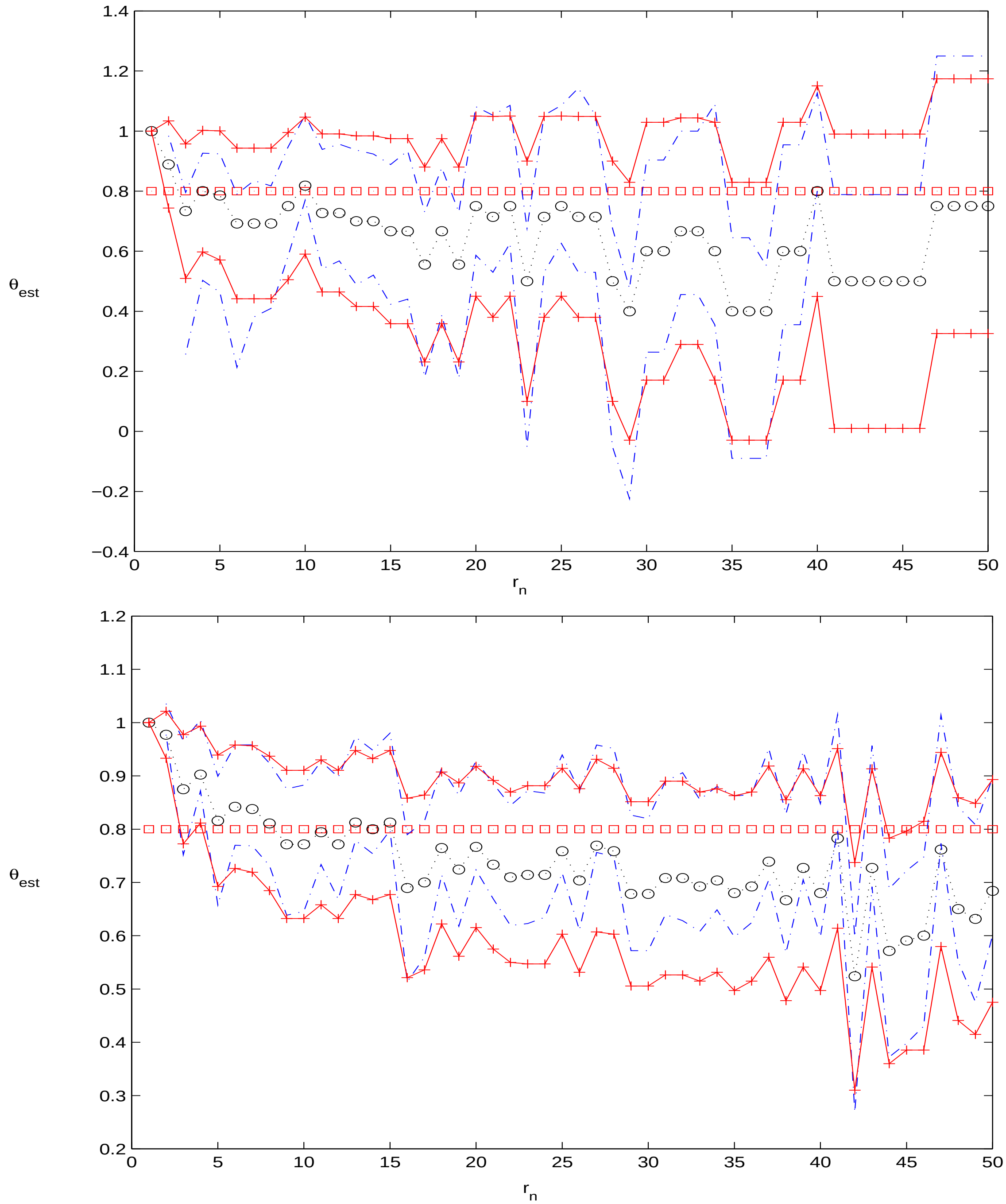


Figure 3.7.1. *Estimated values of θ for the Chernick model with $r = 5$. $\theta = 0.8$ is plotted by $\square$ line. The partitions r_n considered are in the range $[1, 50]$. $\widetilde{\theta}_n$ is represented by $(\cdots)$ and o; the dash line describes the bootstrap confidence interval with $B = 1000$ and $(+-)$ is employed for the asymptotic intervals. The significance level is $\alpha = 0.05$. The sample sizes are $n = 200$ and $n = 1000$ respectively. The threshold is $\hat{u}_n = x_{(n-k)}$ with $k = \sqrt{2n}$.*

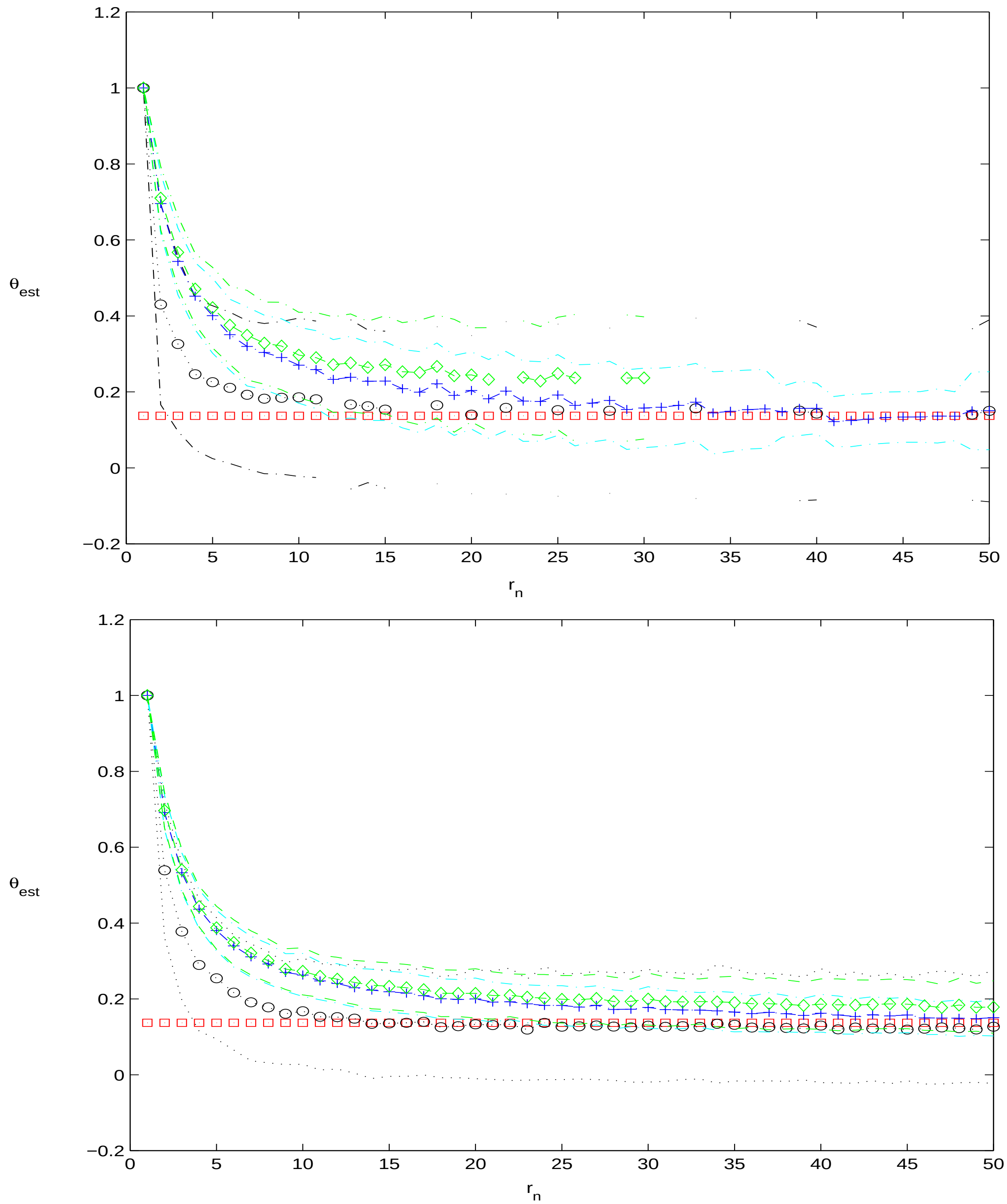
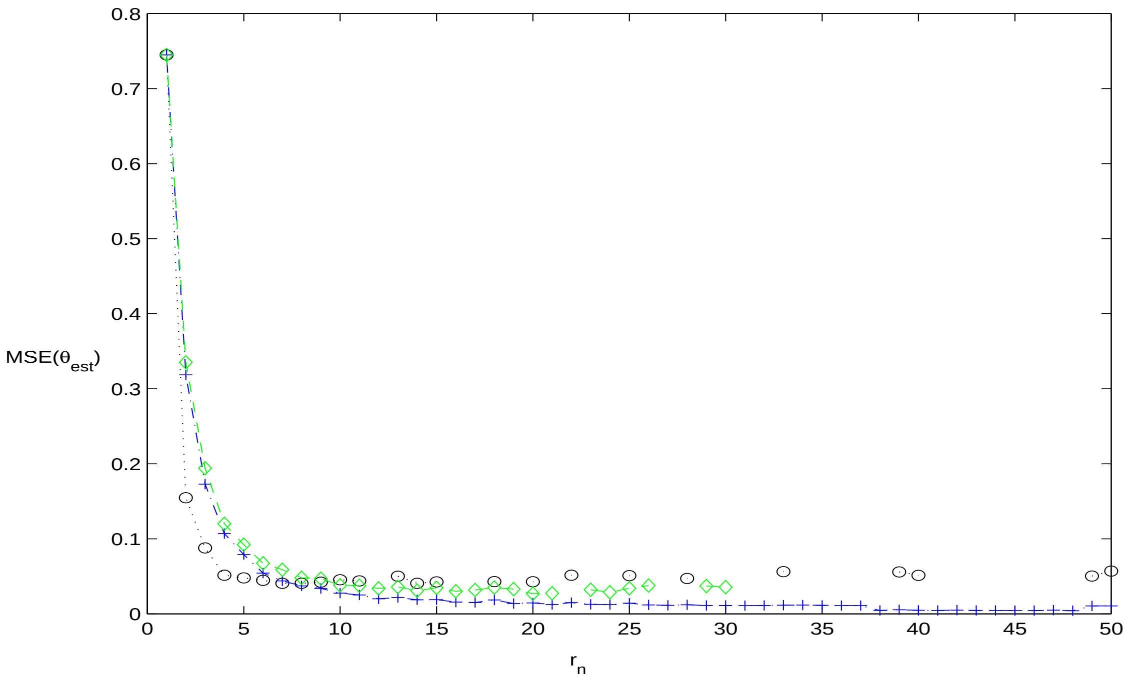


Figure 3.7.2. *Estimated values of θ for the doubly stochastic model with $\psi = 0.9$ and $\eta = 0.7$. The extremal index is $\theta = 0.137$ plotted by $\square$ line. The partitions r_n considered are in the range $[1, 50]$. $\widetilde{\theta}_n$ is represented by $(\cdots)$ and o; the corresponding standard deviation is plotted with $(\cdots)$. The logs method $\hat{\theta}_n^{(1)}$ is represented with $(- - --)$ and $\diamond$. The standard deviation with $(- - --)$. The blocks method $\hat{\theta}_n^{(2)}$ with $(\cdot - \cdot -)$ and $+$, and $(\cdot - \cdot -)$ for the standard deviation. The sample sizes are $n = 200$ and $n = 1000$ respectively. $m = 100$ simulations are used. The threshold sequence is $\hat{u}_n = x_{(n-k)}$ with $k = \sqrt{2n}$.*



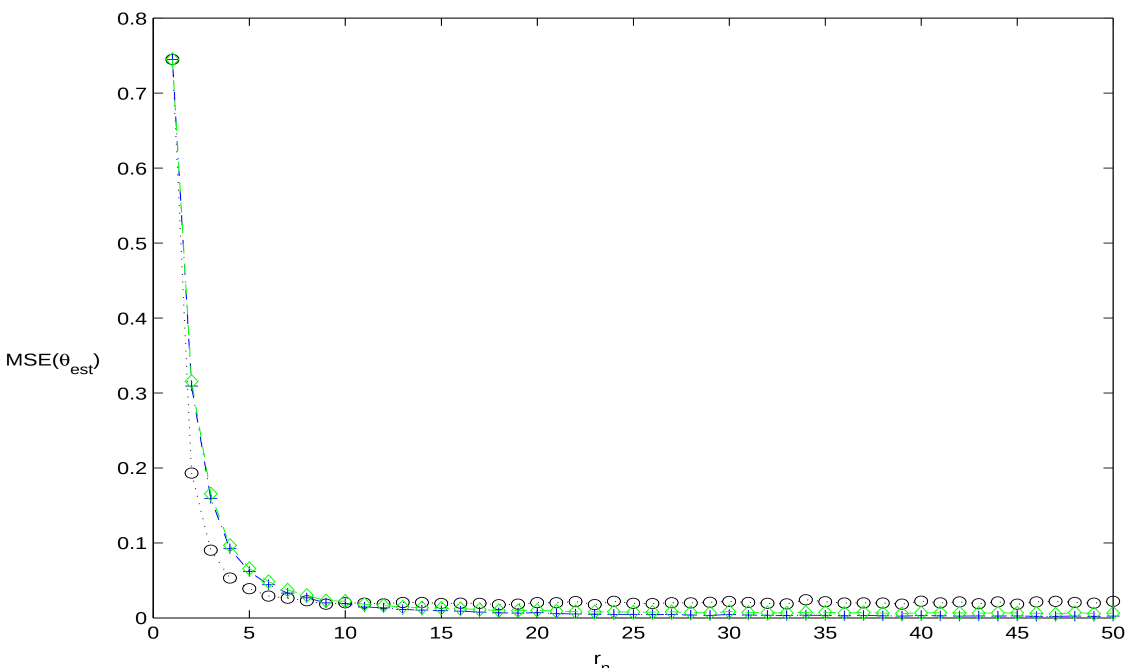


Figure 3.7.3. *Simulated mean square error (MSE) of the estimators of θ for the doubly stochastic model with $\psi = 0.9$ and $\eta = 0.7$. The partitions r_n considered are in $[1, 50]$. $m = 100$ simulations of the model are used. $\widetilde{\theta}_n$ is represented by $(\cdots)$ and o, $\hat{\theta}_n^{(1)}$ with $(- - - -)$ and $\diamond$, and $(\cdot - \cdot -)$ and $+$ for $\hat{\theta}_n^{(2)}$. The sample sizes are $n = 200$ and $n = 1000$ respectively. The threshold sequence is $\hat{u}_n = x_{(n-k)}$ with $k = \sqrt{2n}$.*

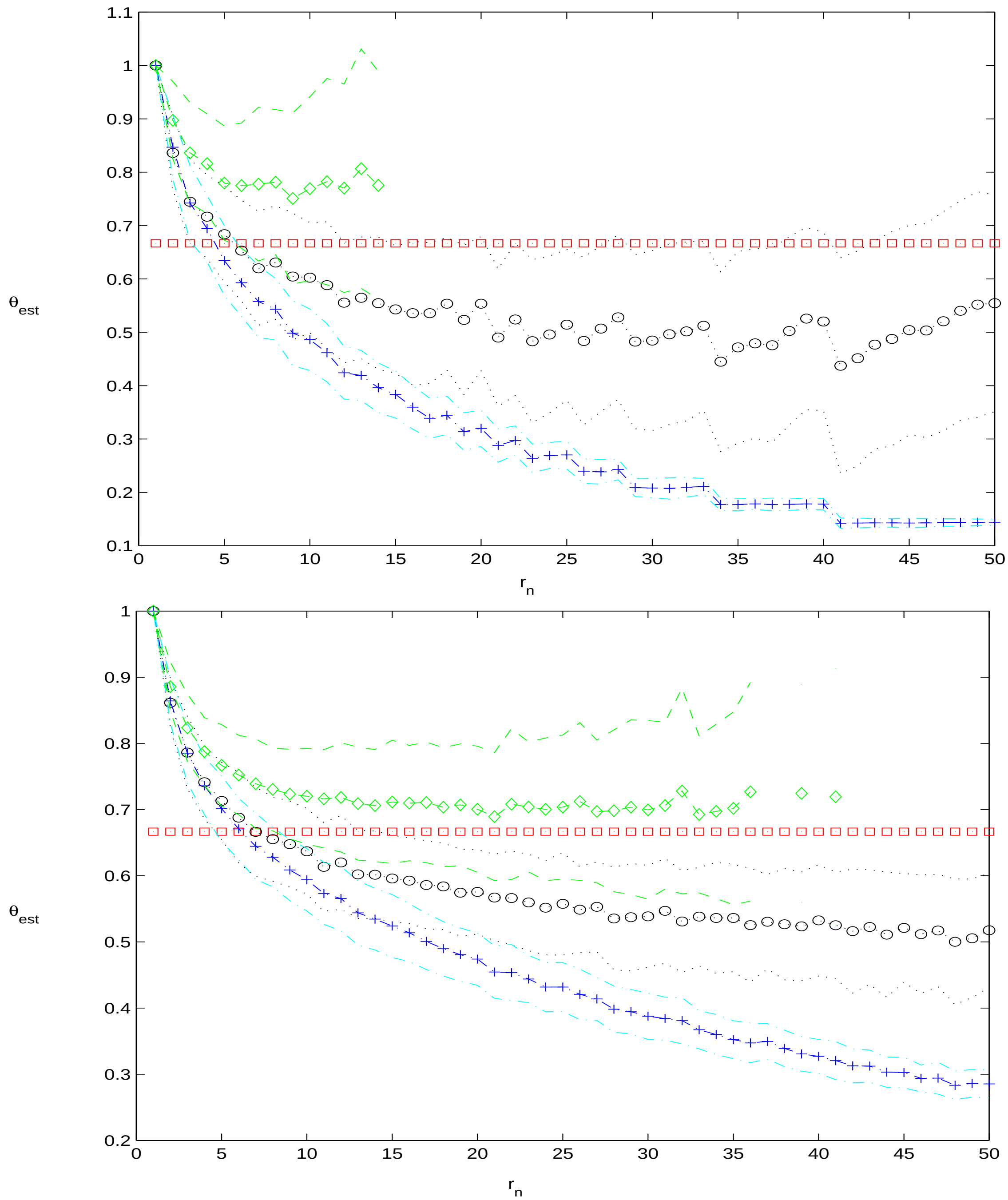


Figure 3.7.4. *Estimated values of the extremal index for the doubly stochastic model with* $\psi = 0.5$ *and* $\eta = 0.5$. $\theta = 0.66$ *is plotted by* $\Box$ *line. The partitions* r_n *considered are in the range* $[1, 50]$. $\widetilde{\theta}_n$ *is represented by* $(\cdots)$ *and o; the corresponding standard deviation is plotted with* $(\cdots)$. *The logs method* $\hat{\theta}_n^{(1)}$ *is represented with* $(- - --)$ *and* $\diamond$. *The standard deviation with* $(- - --)$. *The blocks method* $\hat{\theta}_n^{(2)}$ *with* $(\cdot - \cdot -)$ *and* $+$, *and* $(\cdot - \cdot -)$ *for the standard deviation. The sample sizes are* $n = 200$ *and* $n = 1000$ *respectively.* $m = 100$ *simulations are used. The threshold sequence is* $\hat{u}_n = x_{(n-k)}$ *with* $k = \sqrt{2n}$.

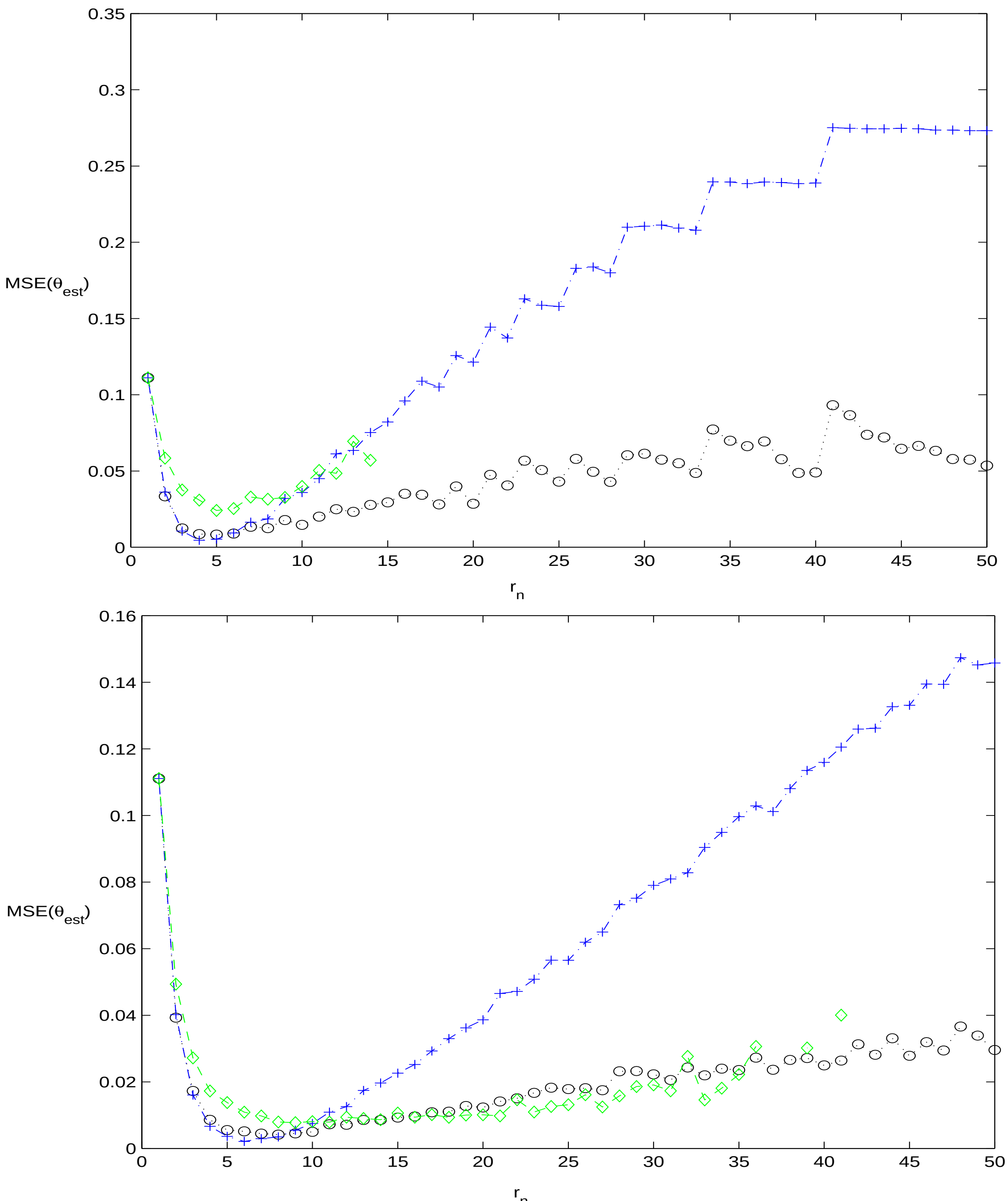


Figure 3.7.5. *Simulated mean square error (MSE) of the estimators of* θ *for the doubly stochastic model with* $\psi = 0.5$ *and* $\eta = 0.5$*. The partitions* r_n *considered are in the range* $[1, 50]$*.* $m = 100$ *simulations of the model are used.* $\widetilde{\theta}_n$ *is represented by* $(\cdots)$ *and* o, $\hat{\theta}_n^{(1)}$ *with* $(- - --)$ *and* $\diamond$*, and* $(\cdot - \cdot -)$ *and* $+$ *for* $\hat{\theta}_n^{(2)}$*. The sample sizes are* $n = 200$ *and* $n = 1000$ *respectively. The threshold sequence is* $\hat{u}_n = x_{(n-k)}$ *with* $k = \sqrt{2n}$*.*

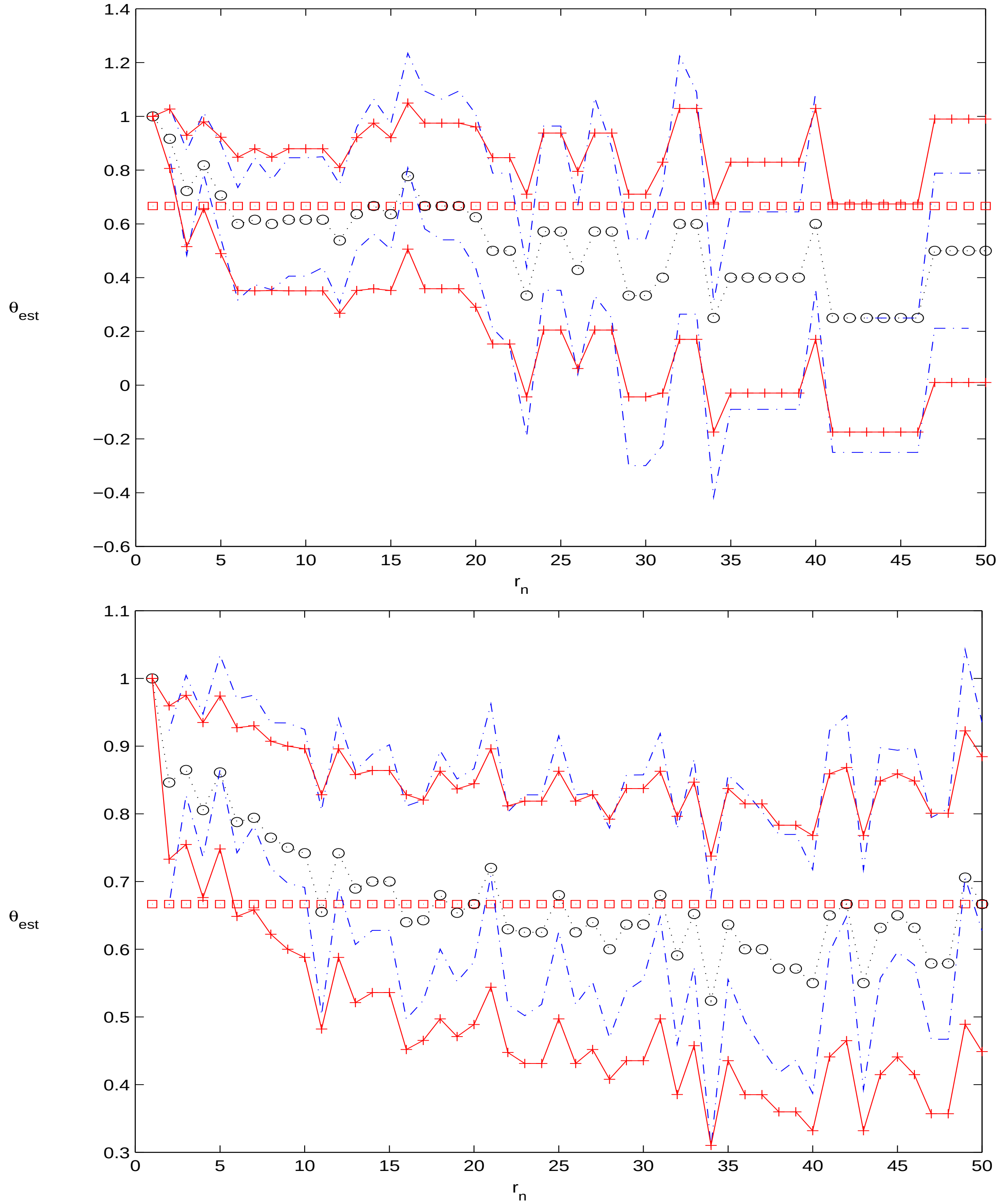


Figure 3.7.6. *Estimated values of the extremal index for the doubly stochastic model with $\psi = 0.5$ and $\eta = 0.5$. The extremal index is $\theta = 0.66$ plotted by $\square$ line. The partitions r_n considered are in the range $[1, 50]$. $\widetilde{\theta}_n$ is represented by $(\cdots)$ and o; the dash line describes the bootstrap confidence interval with $B = 1000$ and $(+-)$ is employed for the asymptotic intervals. The significance level is $\alpha = 0.05$. The sample sizes are $n = 200$ and $n = 1000$ respectively. The threshold sequence is $\hat{u}_n = x_{(n-k)}$ with $k = \sqrt{2n}$.*

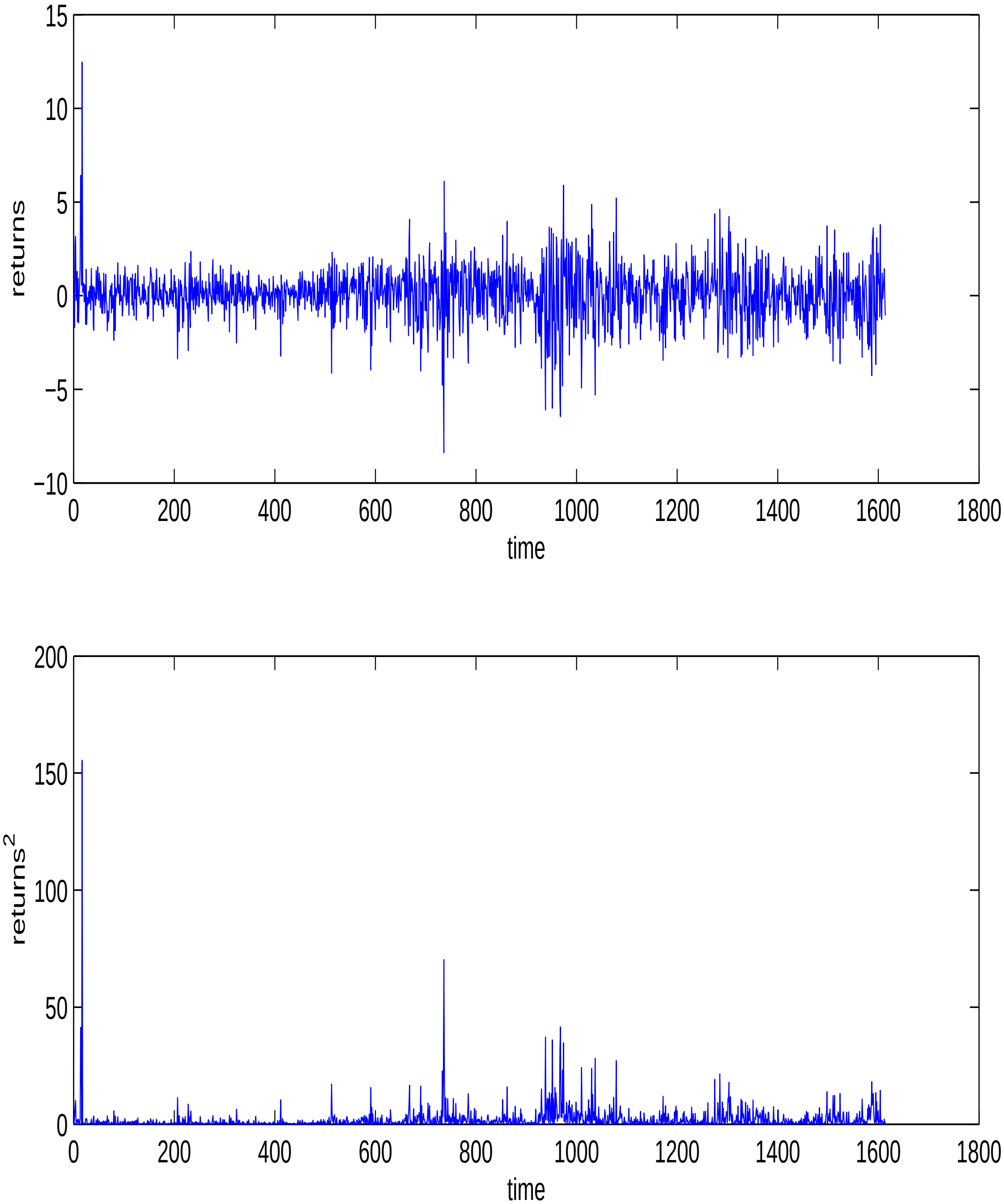


Figure 3.7.7. *DaX Index returns are represented in the upper panel. Squared returns showing the patterns of volatility are plotted in the lower panel. The sample period is* 19/12/1994 − 20/04/2001 *($n = 1614$ observations).*

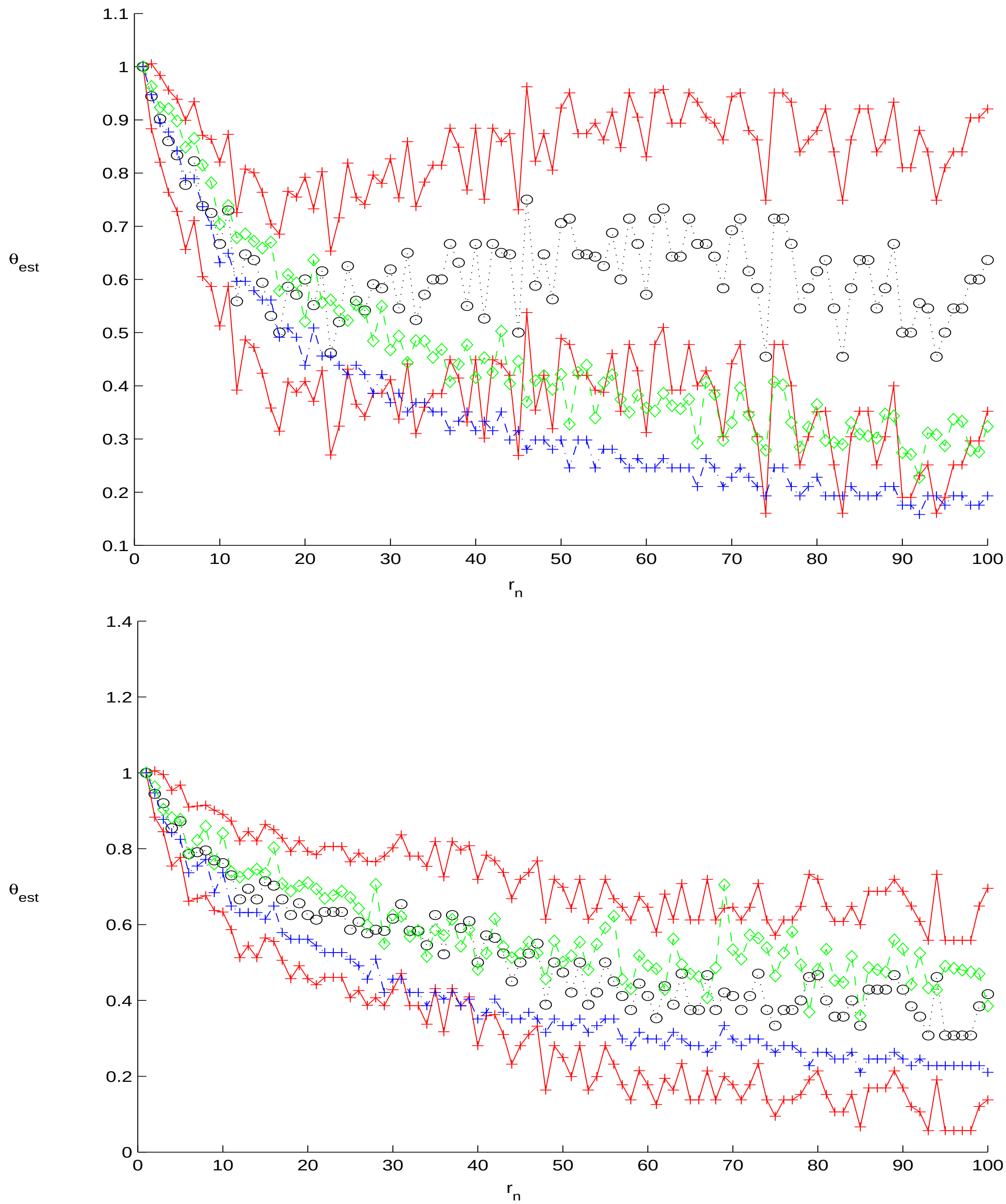


Figure 3.7.8. *Estimated values of θ_{pos} and θ_{neg} for the DaX Index returns over the period 19/12/1994 – 20/04/2001 ($n = 1614$). The upper panel estimates θ_{pos} and the lower panel θ_{neg}. $r_n \in [1, 100]$. $\widetilde{\theta}_n$ is represented by ($\cdots$) and o; ($+-$) describes the asymptotic confidence intervals with $\alpha = 0.05$. $\hat{\theta}_n^{(1)}$ with ($----$) and $\diamond$, and ($\cdot - \cdot -$) and $+$ for $\hat{\theta}_n^{(2)}$. $\hat{u}_{n,pos} = x_{(n-k)}$ and $\hat{u}_{n,neg} = x_{(k)}$ with $k = \sqrt{2n}$ are the corresponding thresholds.*

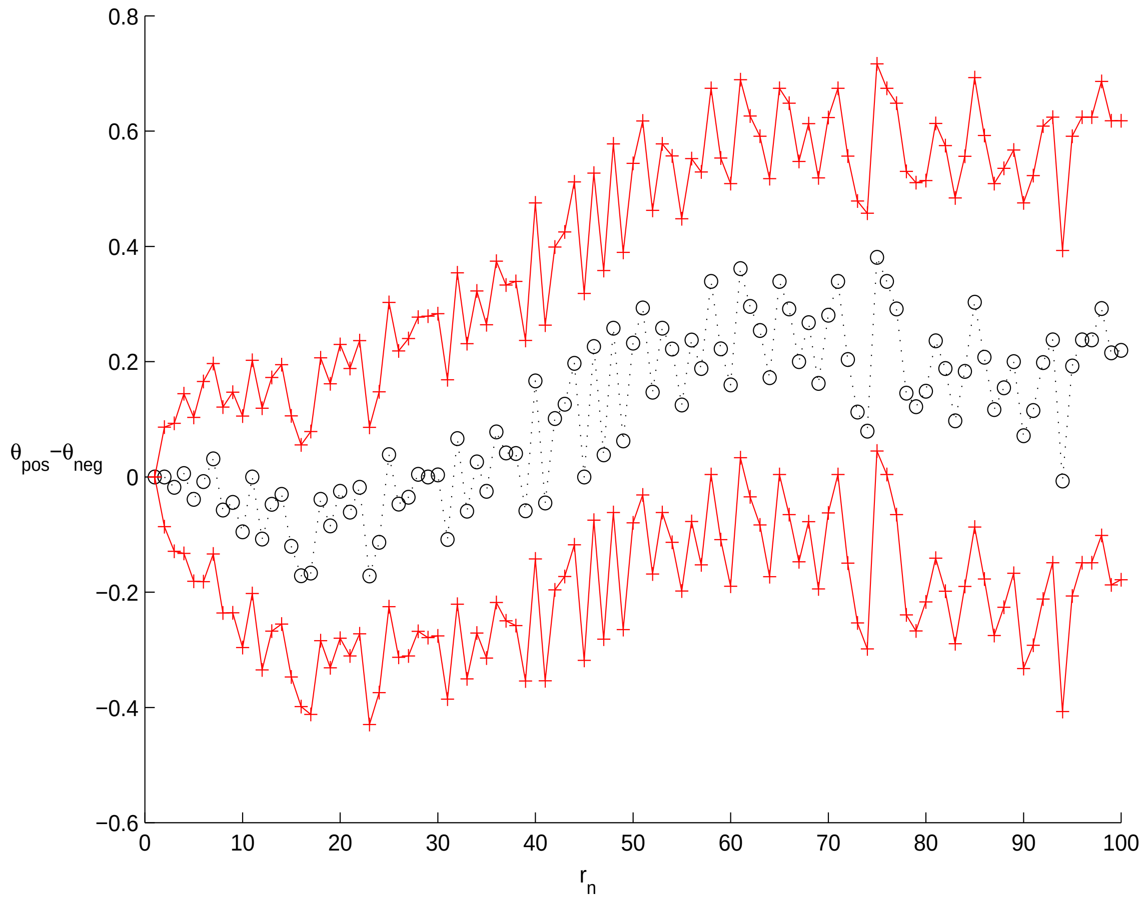


Figure 3.7.9. *Estimated values of $\theta_{pos} - \theta_{neg}$ for the DaX Index returns over the period* $19/12/1994 - 20/04/2001$ *(n=1614).* $r_n \in [1, 100]$. $\widetilde{\theta}_{n,pos} - \widetilde{\theta}_{n,neg}$ *is represented by* $(\cdots)$ *and o;* $(+-)$ *describes the asymptotic confidence intervals with* $\alpha = 0.05$. $\hat{u}_{n,pos} = x_{(n-k)}$ *and* $\hat{u}_{n,neg} = x_{(k)}$ *with* $k = \sqrt{2n}$ *are the corresponding thresholds.*

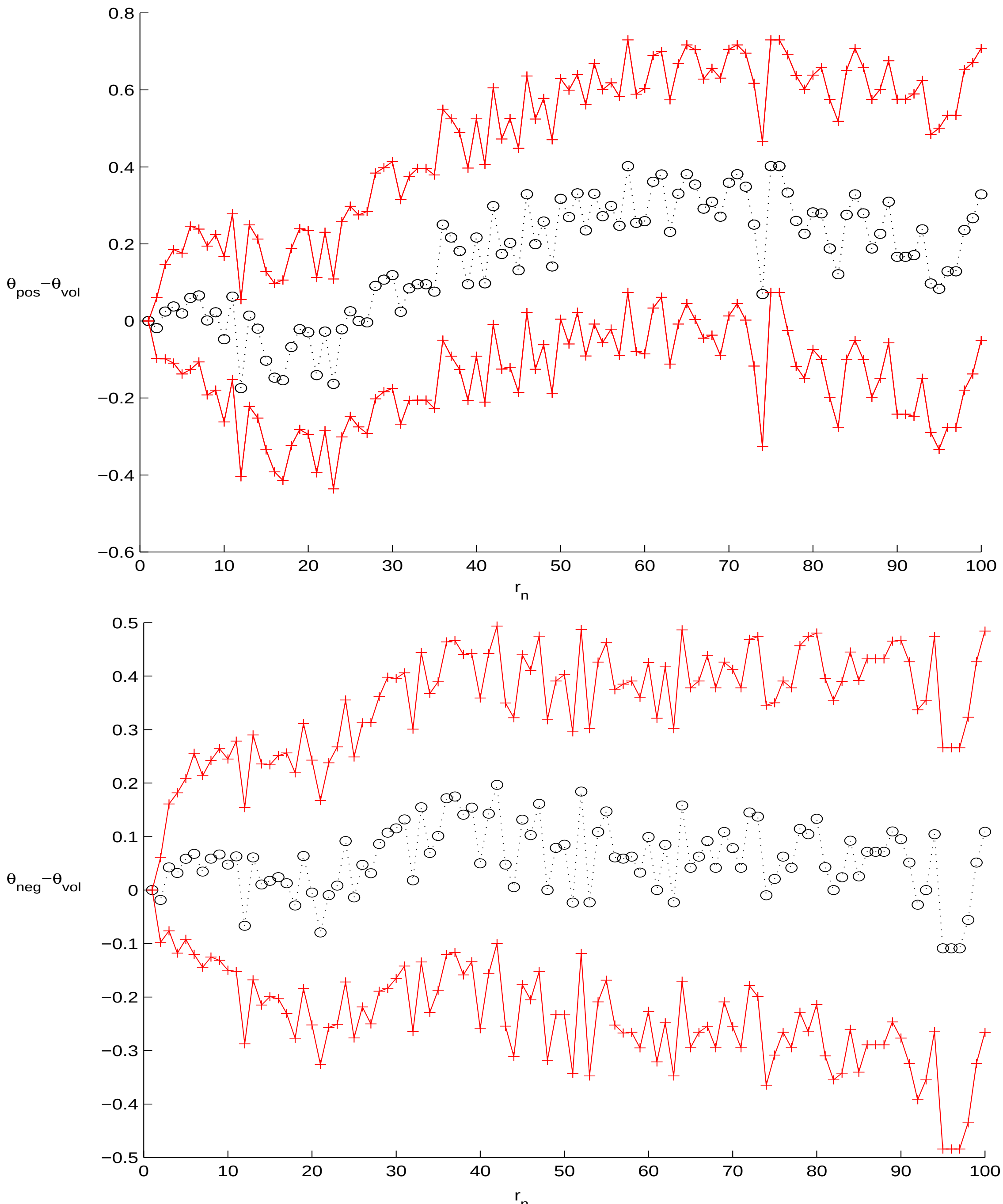


Figure 3.7.10. *Estimated values of $\theta_{pos} - \theta_{vol}$ (upper panel) and $\theta_{neg} - \theta_{vol}$ (lower panel) for the DaX Index returns over the period* 19/12/1994 − 20/04/2001 *(n = 1614). $r_n \in [1, 100]$. $\widetilde{\theta}_{n,pos} - \widetilde{\theta}_{n,vol}$ and $\widetilde{\theta}_{n,neg} - \widetilde{\theta}_{n,vol}$ are represented by ($\cdots$) and o; (+−) describes the asymptotic confidence intervals with $\alpha = 0.05$. $u_{n,pos} = x_{(n-k)}$ is the threshold for the positive exceedances and $u_{n,neg} = x_{(k)}$ for the negative exceedances, with $k = \sqrt{2n}$.*

Chapter 4

Transmission of Risk in Financial Markets: The Contagion effects

None doubts that financial markets are related (interdependent). What is not so clear is whether there exists contagion among them or not, its intensity, and its causal direction. The aim of this work is to define properly the term contagion (different from interdependence) and to present a formal test for its existence, the magnitude of its intensity, and for its direction. Our definition of contagion lies on tail dependence measures and it is made operational through its equivalence with some copula properties. In order to do that, we define a NEW copula, a variant of the Gumbel type, that is sufficiently flexible to describe different patterns of dependence, as well as being able to model asymmetric effects of the analyzed variables (something not allowed with the standard copula models). Finally, we estimate our copula model to test the intensity and the direction of the extreme causality between bonds and stocks markets (in particular, the flight to quality phenomenon) during crises periods. We find evidence of a substitution effect between Dow Jones Corporate Bonds Index with 2 years maturity and Dow Jones Stock Price Index when one of them is through distress periods. On the contrary, if both are going through crises periods a contagion effect is observed. The analysis of the corresponding 30 years maturity bonds with the stock market reflects independent effects of the shocks.

4.1 Introduction

There is common consensus about the concept of crisis, that is, everyone detects a crisis when she is going through one. However, the definitions for this phenomenon are different depending on the features of the economy under study. For example a firm manager concerned about the levels of output may consider that the firm is going through a crisis period if she detects a loss of productivity for certain levels of labor and capital. On the other hand, if most of the firm's business is based on exports abroad the manager will be concerned with sharp appreciations of the local currency against the foreign currency.

These examples raise the issue of finding a general definition of crisis that gathers the different types of crisis regardless the cause. In this way, a naïve and very general definition of crisis in an economy may be given by a threshold that represents a tolerance level. The questions that arise here are how determining this tolerance level and how an exceedance of this threshold affects the tolerance level of other economies or related markets worldwide.

The latter question clearly points out that a crisis is something more than an isolated phenomenon affecting independent markets (financial, credit, currency markets). A crisis in one market is characterized by the collapse not only of that market but by the negative effects produced on other markets. Therefore it seems natural to think of the transmission channels that connect the markets. From an economic viewpoint this involves the analysis of different mechanisms that affect the system: economic fundamentals, market specific shocks, the impact of bad news, or phycological effects (herd behavior). The discussion surges here in the direction and intensity of the dependence between the markets in turmoil periods. There is a large amount of literature concerning these features of dependence. For example Forbes and Rigobon (2001), or Corsetti, Pericoli, Sbracia (2002) where the concepts of interdependence and contagion are analyzed in detail. Regarding the intensity of the dependence, contagion implies that cross market linkages are stronger after a shock to one market, while interdependence implies no significant change in cross market relationships. Regarding the direction, contagion implies that the collapse in one market produces the fall of the other market, whilst interdependence implies that both markets collapse because both are influenced by the same factors.

From an statistical viewpoint, the linkages between markets are usually measured by Pearson correlation. Baig and Goldfajn (1998) compare the correlation between two markets for a pre-crisis and a post-crisis period determined by a shock. They find that there is an increase in cross market correlation after a crisis and therefore there exists a contagion effect. This conditional correlation, however, does not carry the adequate information about an increase in dependence. Forbes and Rigobon (2001) propose an adjusted correlation measure that corrects the problem of conditioning to turmoil periods where cross market correlation is biased upwards because stock market volatility of the conditioning variable (market in crisis) is higher, even if the linkages between the markets remain constant. They find that the cross dependence between the markets is hardly altered after a shock, so there is interdependence but not contagion. Corsetti, Pericoli and Sbracia (2002) find something in the middle, sometimes contagion, sometimes interdependence. They consider that the absence of contagion found in Forbes and Rigobon (2001) can be attributed to pitfalls in their testing procedure.

Correlation, therefore, can lead to misleading results or at least to different interpretations depending on the way of using it. This measure only presents a complete picture of the dependence structure between the markets when their corresponding random variables are jointly gaussian. Under this assumption cross correlations are sufficient to fully describe the dependence structure between the random variables. In this setting multivariate GARCH models are sufficient to describe the dynamics (co-movements) of the vector of random variables. There are many specifications of these models, however a natural specification is given by the extension of the univariate GARCH, that is, the covariances and variances are linear functions of the squares and cross products of the data. Engle and Kroner (1995) propose the vec model that in the first order case is,

$$vec(\Sigma_t) = vec(\Omega) + Avec(X_{t-1}X'_{t-1}) + Bvec(\Sigma_{t-1}),$$

where A, B are $m^2 \times m^2$ matrices with some restrictions, with m the number of random variables. For $m = 2$, $vec(\Sigma_t) = (\sigma^2_{1t}, \sigma_{12t}, \sigma_{21t}, \sigma^2_{2t})$, σ_{it}, $i = 1, 2$ are the conditional volatilities and σ_{12t}, σ_{21t} the conditional covariances at time t. Engle and Kroner (1995) also introduce BEKK models, that in the first order case can be written as

$$\Sigma_t = \Omega + AX_{t-1}X'_{t-1}A' + B\Sigma_{t-1}B',$$

where A, B are $m \times m$ matrices. These models are really complex: the number of parameters to be estimated for the vec model of order 1 is $2m^4$, and for the BEKK model is $2m^2$. In addition, unless the observations are jointly gaussian the cross correlations are not able to fully describe the pattern of multivariate dependence and therefore some dependence is misspecified. Consider for example the asymmetric linkages, corresponding to the left and the right tail, found between most of the financial assets returns. These stylized facts are far from being explained by these models.

Engle (1999) proposes dynamic conditional correlation models (DCC) that extend constant conditional correlation models (CCC) introduced by Bollerslev (1990). The vocation of DCC is to model the structure of dependence between a vector of random variables (m=2) by means of the conditional correlation that is allowed to evolve over time. First the serial dependence of each random variable is individually modelled (GARCH, Stochastic Volatility (SV)), and then the cross dependence between the innovations is modelled by another univariate model (exponential smoothing, GARCH, etc.)

$$\left.\begin{aligned} X_{i,t} &= \epsilon_{it}\sigma_{it}, \quad i = 1, 2, \\ \sigma^2_{i,t} &= \omega_i + \alpha_i X^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}, \end{aligned}\right\} \quad \text{and} \quad \rho^2_t = \omega_o + \alpha_o \epsilon_{1,t-1}\epsilon_{2,t-1} + \beta_o \rho^2_{t-1},$$

with ρ_t the conditional correlation, and ω_i, α_i, β_i, $i = o, 1, 2$ the corresponding parameters of the GARCH processes.

The martingale property is imposed on the vector of innovations, *i.e.* $E[\varepsilon_{it}|\Im_{i,t-1}] = 0$, $i = 1, 2$, with $\Im_{i,t-1}$ the set of information available at $t-1$ for each random variable.

These assumptions do not preclude the case $E[\varepsilon_{it}|\Im_{1,t-1} \bigcup \Im_{2,t-1}] \neq 0$ (Granger causality, Granger (1969)) and the type of specifications for the conditional correlation considered in Engle (1999) are not sufficient to explain the cross linkages between the random variables. Therefore more complex models are called for such that the innovations satisfy $E[\varepsilon_{it}|\Im_{1,t-1} \cup \Im_{2,t-1}] = 0$, $i = 1, 2$. However this assumption does not delivers us from different forms of serial dependence in the innovation vector $(\varepsilon_{1t}, \varepsilon_{2t})$. Instead, we should analyze the whole structure of dependence between the innovations. This is given by the copula function derived from the bivariate distribution $H_t(\varepsilon_{1t}, \varepsilon_{2t})$, and by their conditional counterpart obtained from $H_t(\varepsilon_{1t}, \varepsilon_{2t}|\Im_{1,t-1} \cup \Im_{2,t-1})$, see Patton (2001) or Granger, Terasvirta and Patton (2002).

The definition of copula is due to Sklar (1959). This function provides the complete structure of dependence between the random variables after taking into account the corresponding marginal distributions. In particular, the model introduced in Engle (1999) may be considered a gaussian copula where the dynamics of the dependence are given by the conditional correlation. The set of available copulas is endless providing different alternatives suiting to the problem at hand. Some examples are given by the gaussian copula used by Longin and Solnik (2001) to describe the dependence in financial asset returns, Student's t copulas (Mashal and Zeevi, (2002)) that suit better to the tails of these sequences, Joe-Clayton copula in Joe (1997) or its variation, the symmetrized Joe-Clayton copula introduced in Patton (2001) for the dependence between exchange rates series.

It is also interesting to analyze the links between the vector of random variables in the tail regions. Its joint distribution function in the tail region is derived from the multivariate extreme value theory, see Resnick (1987). Applications of multivariate extreme value distributions to examples concerning the tail regions appear in Ledford and Tawn (1996) or in de Haan and de Ronde (1998). The analysis of the dependence in the extremes provides an interesting alternative to correlation for measuring the strength of the linkages between the random variables as they become more extreme (differences between interdependence and contagion).

The vocation of this chapter is modelling the dependence found between the random variables that represent different economies and financial markets. This dependence is divided in two classes regarding the origin. First, the links due to economic fundamentals (rational dependence) and second, the co-movements of the corresponding innovations (irrational dependence). Our focus lies on the latter form of dependence and the concepts of interdependence and contagion. In order to model this form of dependence (cross dependence in the innovations sequences) we introduce an innovative copula function derived from the extreme value theory that incorporates sufficient flexibility to describe different patterns of dependence, in particular asymmetric effects between the variables not reflected by standard copulas. Furthermore, the concepts of interdependence and contagion are revisited and the definitions proposed in the literature are adapted to be expressed as tail dependence measures, and in turn properties of the copula functions involving the tails of the marginal distributions. Finally, the intention of the authors is to apply this methodology to test the flight to quality phenomenon, that is, outflows of capital from the stocks markets to the bonds markets when the first ones are facing crises periods.

This chapter is structured as follows. Section 2 introduces the copula function derived from the multivariate extreme value theory. The next section proposes tail dependence measures as an alternative to correlation; these measures are used to define contagion and interdependence. The cases of asymptotic dependence and independence are also studied. Finally the section studies the statistical aspects of the model, and provides a test for the existence of these effects. In Section 4, this innovative copula function as well as the new dependence measures are applied to analyze the dependence structure between bonds and assets (flight to quality phenomenon). Section 5 presents the conclusions.

4.2 The model

Consider the model

$$\left.\begin{aligned} X_{1,t} &= g_1(X_{1,t-1},\ldots,X_{m,t-1}) + \varepsilon_{1,t}, \\ &\ldots \quad \ldots \quad \ldots \\ X_{m,t} &= g_m(X_{1,t-1},\ldots,X_{m,t-1}) + \varepsilon_{m,t}, \end{aligned}\right\} \tag{4.1}$$

and assume that $(\varepsilon_{1,1},\ldots,\varepsilon_{m,1}),\ldots,(\varepsilon_{1,t},\ldots,\varepsilon_{m,t})$ are independent vectors, that is, the multivariate dependence between the innovations is given by $H_t((\varepsilon_1,\ldots,\varepsilon_m)|\Im_{t-1})$, with $\Im_{t-1} = \Im_{X_1,t-1} \cup \ldots \cup \Im_{X_m,t-1}$. Note that the structure of dependence is time varying though the marginal distributions of the observations are independent of time. Otherwise if the innovations satisfied the martingale property, the marginal distributions would not be free from the time index, that is, the joint distribution function would take the form

$$H_t((\varepsilon_{1,t}|\Im_{t-1},\ldots,\varepsilon_{m,t}|\Im_{t-1})|\Im_{t-1}),$$

with $\varepsilon_{i,t}|\Im_{t-1}$, $i=1,\ldots,m$ the conditional random variables.

Both distribution functions, however, give rise to the type of conditional copulas introduced in Patton (2001) where the dynamics of the joint distribution function is driven by a parameter that is time varying. Instead, for appropriate functions $g_1,\ldots,g_m$ we propose a multivariate distribution function $H(\varepsilon_1,\ldots,\varepsilon_m)$ time invariant motivated by the dependence found between the vector of maxima of the corresponding random variables.

4.2.1 The structure of dependence: The copula function

This section studies the dependence structure between m random variables via the copula function. The concept of copula is due to Sklar (1959) and refers to the class of multivariate distribution functions supported in the unit cube with uniform margins.

Definition 4.2.1. *A function $C:[0,1]^m \to [0,1]$ is a m-dimensional copula if it satisfies the following properties:*

- *For all $u_i \in [0,1]$, $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$.*
- *For all $u \in [0,1]^m$, $C(u_1,\ldots,u_m) = 0$ if at least one of the coordinates is zero.*
- *The volume of every box contained in $[0,1]^m$ is non-negative, i.e., $V_C([u_1,\ldots,u_m]\times[v_1,\ldots,v_m])$ is non-negative. For $m=2$, $V_C([u_1,u_2]\times[v_1,v_2]) = C(u_2,v_2)-C(u_1,v_2)-C(u_2,v_1)+C(u_1,v_1) \geq 0$ for $0 \leq u_i, v_i \leq 1$.*

The copula $C(u_1, \ldots, u_m)$ is the joint distribution function of the probability integral transforms of each of the variables $X_1, \ldots, X_m$ with respect to the marginal distributions $F_1, \ldots, F_m$. It may be seen as the component of the multivariate distribution function of a vector of random variables that captures the dependence structure.

Theorem 4.2.1. *(Sklar's theorem): Given a m-dimensional distribution function H with continuous marginal distributions $F_1, \ldots, F_m$, then there exists a unique copula $C : [0,1]^m \rightarrow [0,1]$ such that*

$$H(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)), \quad \forall x_1, \ldots, x_m \in \mathbb{R} \cup \{\infty\}. \tag{4.2}$$

Conversely, if $C(u_1, \ldots, u_m)$ is a m-dimensional distribution function with uniform margins, and $F_1, \ldots, F_m$ are continuous univariate distribution functions for the random variables $X_1, \ldots, X_m$, then the function H defined in (4.2) is a m-dimensional distribution function with margins $F_1, \ldots, F_m$.

It is immediate to see that if we have a model for the joint distribution of the m random variables and the marginal distributions of the X_i are continuous, the complete dependence structure of the corresponding variables is known,

$$C(u_1, \ldots, u_m) = H(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)), \tag{4.3}$$

with $F_i^{-1}(u) = inf\{x \in \mathbb{R} | F_i(x) \geq u\}$, for all $0 \leq u \leq 1$.

This measure of dependence extends the notions of linear correlation (Pearson) and rank correlation (Spearman). More important, it overcomes the typical problems of these scalar measures. Embrechts, McNeil and Straumann (1999) provides an excellent review about the properties and problems of these dependence measures.

It is shown that under very general conditions on the marginal distribution functions the dependence structure of any multivariate distribution is described by the copula function. In particular this interesting result is found for the joint distribution of the maxima of a vector of random variables. Moreover, there exists a copula function that drives the dependence in the extremes whose expression is derived from the extreme value theory.

Consider $M_n = (M_{n1}, \ldots, M_{nm})$ the vector of componentwise maxima, with components $M_{ni} = \max\{X_{1i}, \ldots, X_{ni}\}$, and the vector of sequences $a_n = (a_{n1}, \ldots, a_{nm})$ with each $a_{ni} > 0$, and $b_n = (b_{n1}, \ldots, b_{nm})$. Under some smoothness condition in the tail of F_i, Leadbetter, Lindgren and Rootzén (1983) show that

$$\lim_{n\to\infty} F_i^n(a_{ni}x_i + b_{ni}) = G_i(x_i), \quad i = 1, \ldots, m, \tag{4.4}$$

where $G_i(x_i)$ is an extreme value distribution of one of the three possible types, Gumbel, Weibull and Fréchet. The distribution F_i is said to belong to the maximum domain of attraction of G_i, see also Embrechts, Klüppelberg and Mikosch (1997). Denote the distribution of the multivariate maximum by

$$H^n(a_{n1}x_1 + b_{n1}, \ldots, a_{nm}x_m + b_{nm}) = P\{a_{ni}^{-1}(M_{ni} - b_{ni}) \leq x_i, i = 1, \ldots, m\}, \tag{4.5}$$

where $H(x_1, \ldots, x_m) = P\{X_1 \leq x_1, \ldots, X_m \leq x_m\}$. The core result of the multivariate extreme value theory is that (4.4) may be extended to

$$\lim_{n\to\infty} H^n(a_{n1}x_1 + b_{n1}, \ldots, a_{nm}x_m + b_{nm}) = G(x_1, \ldots, x_m), \tag{4.6}$$

with G a non degenerate multivariate extreme value distribution (*mevdf*). The class of these particular distributions is precisely the class of max-stable distributions (Resnick (1987), proposition 5.9). These distributions are defined by this property

$$G^t(tx_1, \ldots, tx_m) = G(\alpha_1 x_1 + \beta_1, \ldots, \alpha_m x_m + \beta_m), \tag{4.7}$$

for every $t > 0$, and some constants $\alpha_i > 0$ and β_i.

The marginal distribution functions of G are the univariate extreme value distributions

$G_i(x_i)$. By Sklar's theorem, (4.6) may be written as

$$\lim_{n\to\infty} H^n(a_{n1}x_1 + b_{n1}, \ldots, a_{nm}x_m + b_{nm}) = C(G_1(x_1), \ldots, G_m(x_m)), \tag{4.8}$$

with C a copula function.

It can be seen under some simple algebra that C also describes the dependence structure of the largest observations. Our aim in the following lines is to derive a suitable analytical expression for this copula function. In order to do this, the marginal distributions are transformed to obtain identical and parameter free versions of these univariate distributions, in particular Fréchet distributions of the type $\Psi_\alpha(z) = \exp(-z^{-\alpha})$ with $\alpha = 1$.

Let $Z_i = 1/log\frac{1}{F_i(X)}$ be such transformation, and denote $P\{Z_i \leq z\} = F_i^*(z)$ with $z = 1/log\frac{1}{F_i(a_{ni}x+b_{ni})}$. This distribution satisfies these interesting properties: $F_i^*(z) = \Psi_1(z)$, $F_i^*(z) = F_i(a_{ni}x + b_{ni})$ and $F_i^{*n}(nz) = \Psi_1(z)$. Note that these conditions imply $F_i^{*n}(nz) = F_i^*(z)$ and

$$\lim_{n\to\infty} H^{*n}(nz_1, \ldots, nz_m) = C(\Psi_1(z_1), \ldots, \Psi_1(z_m)), \tag{4.9}$$

with $H^*(z_1, \ldots, z_m) = H(a_{n1}x_1+b_{n1}, \ldots, a_{nm}x_m+b_{nm})$. This condition holds for any vector $(z_1, \ldots, z_m)$ in $[z_{o1}, \infty) \times \ldots \times [z_{om}, \infty)$, with $(z_{o1}, \ldots, z_{om})$ a threshold vector. The function C is called extreme copula because satisfies this property,

$$C^t(\Psi_1(z_1), \ldots, \Psi_1(z_m)) = C(\Psi_1^t(z_1), \ldots, \Psi_1^t(z_m)), \quad t > 0, \tag{4.10}$$

where the margins are extreme value distributions. The proof immediately follows from (4.7). This condition entails an invariance property given by the logs of the corresponding distributions, that is,

$$t\,log\,C(\Psi_1(tz_1), \ldots, \Psi_1(tz_m)) = log\,C(\Psi_1(z_1), \ldots, \Psi_1(z_m)).$$

Then, for n and $(z_1, \ldots, z_m)$ sufficiently high,

$$\lim_{n\to\infty} n\,(1 - H^*(nz_1, \ldots, nz_m)) = -log\, C(\Psi_1(z_1), \ldots, \Psi_1(z_m)), \tag{4.11}$$

and

$$\lim_{n\to\infty} \frac{H^*(nz_1, \ldots, nz_m)}{1 + log\, C(\Psi_1(nz_1), \ldots, \Psi_1(nz_m))} = 1. \tag{4.12}$$

Other interesting result derived from the invariance property is

$$P\left\{Z_1 \le nz_1, \ldots, Z_m \le nz_m | \bigcup_{i=1}^{m} Z_i > nz_{0i}\right\} = P\left\{Z_1 \le z_1, \ldots, Z_m \le z_m | \bigcup_{i=1}^{m} Z_i > z_{0i}\right\}.$$

The left term in (4.11) may be considered as a sequence of measures that converge to a constant given the vector $(z_1, \ldots, z_m)$, see Resnick (1987) or de Haan and de Ronde (1998) for different transformations of the marginal distributions. Expression (4.12) provides the joint distribution function of the largest observations, that is, for n sufficiently high the denominator may be approximated by the copula function C. Therefore

$$P\{Z_1 \le z_1, \ldots, Z_m \le z_m\} = C(\Psi_1(z_1), \ldots, \Psi_1(z_m)), \tag{4.13}$$

for the vector $(z_1, \ldots, z_m)$ sufficiently high.

The latter expression is promising in the sense that C is a good approximation of the dependence structure in the largest observations. However, the challenge of choosing a suitable threshold vector that determines the region satisfying condition (4.11) still remains.

On the other hand the invariance property implies that the copula C must be of exponential type. There are different characterizations of this distribution. A general expression for $m = 2$ is given in the form of the Pickands representation, (Pickands, 1981) that is,

$$C(u_1, u_2) = \exp^{D(t) log(u_1 u_2)}, \tag{4.14}$$

where $u_1 = \Psi_1(z_1)$, $u_2 = \Psi_1(z_2)$, $t = \frac{log(u_1)}{log(u_1 u_2)}$, and $D(t)$ is a convex function on $[0, 1]$ such that $\max(t, 1-t) \leq D(t) \leq 1$ for all $0 \leq t \leq 1$. This family of distributions is included in the class of Archimedean copulas (Nelsen 1999, chapter 4). The dependence in these copulas is driven by a single variable t for $m = 2$. The Gumbel-Hougaard family is within this class of distributions and satisfies the invariance property. It is represented by

$$C_G(u_1, u_2; \theta) = \exp^{-[(-log\ u_1)^\theta + (-log\ u_2)^\theta]^{1/\theta}}, \quad \theta \geq 1. \tag{4.15}$$

This distribution function is usually known as Gumbel bivariate logistic copula. The main problem that arises if we assume that C is modelled by a Gumbel distribution C_G in (4.13) is the choice of the threshold. Condition (4.11) may be violated for low thresholds where the extreme value theory is not a reliable technique. Other drawback of modelling C as C_G is the asymmetry, the random variables modelled by the Gumbel copula are exchangeable and hence it is not possible to quantify different contributions of the corresponding random variables. In order to account for this asymmetric dependence we propose a version of C_G able to describe these effects. This function is denoted by $\widetilde{C}_G(u_1, u_2; \Theta)$, with $\Theta = \{\theta, \gamma, \eta\}$, and takes the following expression

$$\widetilde{C}_G(u_1, u_2; \Theta) = \exp^{-D(u_1, u_2; \gamma, \eta)[(-log\ u_1)^\theta + (-log\ u_2)^\theta]^{1/\theta}}, \tag{4.16}$$

with

$$D(u_1, u_2; \gamma, \eta) = \exp^{\gamma(1-u_1)(1-u_2)^\eta}, \quad \gamma \geq 0, \quad \eta > 0. \tag{4.17}$$

Theorem 4.2.2. *The function $\widetilde{C}_G : [0,1] \times [0,1] \to [0,1]$ defined in (4.16) and (4.17) is a copula function if the parameters in Θ satisfy that $\widetilde{c}_G(u_1, u_2; \Theta) > 0$, $\forall\, (u_1, u_2) \in [0,1] \times [0,1]$, with $\widetilde{c}_G(u_1, u_2; \Theta) = \frac{\delta^2 \widetilde{C}_G(u_1, u_2; \Theta)}{du_1 du_2}$ the density function of the copula $\widetilde{C}_G$.*

Proof.- Let us denote $A(u_1, u_2; \theta) = [(-log\ u_1)^\theta + (-log\ u_2)^\theta]^{1/\theta}$. The conditions related to the contour of $\widetilde{C}_G$ immediately follow from the contour properties of the functions $D(u_1, u_2; \gamma, \eta)$ and $A(u_1, u_2; \theta)$. The proof that $\widetilde{C}_G$ is 2-increasing involves more algebra. Consider $V_{\widetilde{c}_G}([u_{o1}, u_{11}] \times [u_{o2}, u_{12}]) = \widetilde{C}_G(u_{11}, u_{12}; \Theta) - \widetilde{C}_G(u_{11}, u_{o2}; \Theta) - \widetilde{C}_G(u_{o1}, u_{12}; \Theta) +$

$\widetilde{C}_G(u_{o1}, u_{o2}; \Theta)$, and define $V'(u_1) = \widetilde{C}_G(u_1, u_{12}; \Theta) - \widetilde{C}_G(u_1, u_{o2}; \Theta)$. Then, $V_{\widetilde{c}_G}([u_{o1}, u_{11}] \times [u_{o2}, u_{12}]) = V'(u_{11}) - V'(u_{o1})$. Note that $V'(u_1) \geq 0$, $\forall\, u_1 \in [0,1]$, with $u_{o2} < u_{12}$. This function can be written as

$$V'(u_1) = \exp^{-D(u_1,u_{o2};\gamma,\eta)A(u_1,u_{o2};\theta)} \left[\exp^{-(D(u_1,u_{12};\gamma,\eta)A(u_1,u_{12};\theta) - D(u_1,u_{o2};\gamma,\eta)A(u_1,u_{o2};\theta))} - 1\right],$$

that is greater than 0 if and only if $D(u_1, u_2; \gamma, \eta)A(u_1, u_2; \theta)$ is decreasing in u_2. The only condition that remains to see is that $V'(u_1)$ is nondecreasing. This condition will hold if the function $\frac{\delta \widetilde{C}_G(u_1,u_2;\Theta)}{du_1}$ is nondecreasing in u_2, that amounts to see if $\widetilde{c}_G(u_1, u_2; \Theta) > 0$, $\forall\, (u_1, u_2) \in [0,1] \times [0,1]$. $\square$

The choice of the threshold in (4.13) is overcome by adding the function $D(u_1, u_2; \gamma, \eta)$. This function by means of the pair $(u_1,\, u_2)$ and the parameter γ measures the sensitivity of the dependence structure to departures from the invariance property. In other words, either the margins are further in the right tail $(u_1,\, u_2 \to 1)$ or $\gamma \cong 0$ the copula function $\widetilde{C}_G$ is closer to C_G and the invariance property holds. In this way the joint distribution function for the entire range of the random variables Z_1, Z_2 is

$$P\left\{Z_1 \leq z_1, Z_2 \leq z_2\right\} = \widetilde{C}_G(\Psi_1(z_1), \Psi_1(z_2)), \tag{4.18}$$

where $z_i = 1/log\, \frac{1}{F_i(x_i)}$ in this case. This distribution function is driven by the parameters θ, γ, η. The constant γ assesses the extent of the invariance property. The parameter θ describes the level of asymptotic tail dependence between the random variables. The case of perfect independence is covered by $\theta = 1$, $\gamma = 0$. Finally η measures the level of asymmetry or exchangeability of the variables.

The following list enumerates the most outstanding advantages of our copula function $\widetilde{C}_G$.

1. This copula function is derived from the multivariate extreme value theory, in contrast to ad-hoc choices to model the dependence structure.

2. The function $D(u_1, u_{12}; \gamma, \eta)$ and in particular the parameter γ extend the results of the multivariate extreme value theory about the distribution of the largest observations to the entire range of the random variables.

3. $\widetilde{C}_G$ is able to explain asymmetric effects of the variables for $\eta \neq 1$. It may be considered as an alternative to the asymmetric logistic model in Tawn (1988).

4. This copula function is sufficiently flexible to describe different forms of dependence and asymptotic dependence, as will be shown below.

4.3 Contagion: types and definitions

Linear measures of dependence are not sufficient to describe the dependence patterns between a vector of random variables. The popular Pearson correlation has a number of pitfalls, see Embrechts, McNeil and Straumann (1999). Some of them are that a a zero correlation does not imply independence if the marginal distributions are not elliptical, and second, correlation is not invariant under transformations of the random variables. Spearman correlation (rank correlation) for example solves the latter, however, it also fails to give a measure of independence far from the elliptical world.

In the bivariate setting natural measures of dependence different from the traditional correlation are given by the dependence in the tails. Ledford and Tawn (1997) and Coles, Heffernan and Tawn (1999) define the asymptotic tail dependence measure $\aleph$,

$$\aleph = \lim_{t \to \infty} P\{Z_2 > t | Z_1 > t\}. \tag{4.19}$$

This measure takes the zero value if the random variables are asymptotically independent. There are two classes of extreme value dependence, asymptotic dependence and asymptotic independence. Both forms of dependence permit dependence for moderately large values of the variables, however the likelihood of joint extreme events under asymptotic independence converges to 0 as the events become more extreme. Loosely speaking, the probability of one variable being extreme given the other is extreme is 0 in the limit. The copula $\widetilde{C}_G$ supports both types of asymptotic dependence. It can be seen that $\aleph_{\widetilde{C}_G} = 2 - 2^{1/\theta}$, that reflects asymptotic independence for $\theta = 1$ and asymptotic dependence otherwise.

The definition in (4.19) can be extended to the entire range of the random variables. Lehman (1966) defined two random variables Z_1, Z_2 as positively quadrant dependent (PQD) if for all $(z_1, z_2) \in \mathbb{R}^2$,

$$P\{Z_1 > z_1, Z_2 > z_2\} \geq P\{Z_1 > z_1\}P\{Z_2 > z_2\}, \tag{4.20}$$

or equivalently if

$$P\{Z_1 \leq z_1, Z_2 \leq z_2\} \geq P\{Z_1 \leq z_1\}P\{Z_2 \leq z_2\}. \tag{4.21}$$

In the same way negative quadrant dependence (NQD) is defined reversing the inequalities in both expressions.

Definition 4.3.1. *We say that two random variables are interdependent if they are PQD. In consequence interdependence is characterized by joint movements in the same direction (co-movements) of the corresponding random variables.*

If Z_1 and Z_2 are NQD a large value in one random variable is corresponded by a value of the same magnitude in the opposite direction for the other variable. Economically, interdependence means that links in turmoil periods (tails of the distributions) are only consequence of the same linkages between the markets found in still periods.

In the case that the random variables are continuous these definitions are a property of the copula. From elementary probability theory

$$P\{Z_1 > z_1, Z_2 > z_2\} = \widetilde{C}_G(u_1, u_2) - (u_1 + u_2) + 1, \tag{4.22}$$

with $u_i = \Psi_1(z_i)$, $i = 1, 2$.

Define the function $g(u_1, u_2)$ as the difference between the probabilities in (4.20) in terms of the copula function,

$$g(u_1, u_2) = \widetilde{C}_G(u_1, u_2) - u_1 u_2. \tag{4.23}$$

If this function is positive for all $(u_1, u_2) \in [0, 1] \times [0, 1]$ the former definitions for cross dependence apply, that is, Z_1 and Z_2 are *PQD*.

The function $g(u_1, u_2)$ itself is not sufficient to determine the strength of the links between the variables. A stronger condition is required to measure the amount of dependence for different values of the random variables. This condition is tail monotonicity, that is, the function (4.23) is either nonincreasing or nondecreasing in its arguments. In particular, increasing tail monotonicity for the function $P\{Z_1 > z_1, Z_2 > z_2\} - P\{Z_1 > z_1\}P\{Z_2 > z_2\}$ characterizes the existence of contagion in the upper tails between the random variables. Thus, contagion in this context can be defined as a significant increase in the intensity of the dependence between the variables Z_1, Z_2 when these take on extreme values.

Definition 4.3.2. *Suppose Z_1, Z_2 with common Fréchet distribution Ψ_1 and consider z a threshold that determines the extremes in the right tail of both random variables. Then, there exists a contagion effect between Z_1 and Z_2 if $g(u_1, u_2)$ is an increasing function for both random variables, and for $u_1, u_2 \geq u$ with $u = \Psi_1(z)$.*

On the other hand contagion in intensity for the lower tails is characterized by decreasing tail monotonicity for the function $P\{Z_1 \leq z_1, Z_2 \leq z_2\} - P\{Z_1 \leq z_1\}P\{Z_2 \leq z_2\}$. In terms of copulas the conditions in definition 4.3.2 for contagion amount to these properties,

$$h_1(u_1, u_2) = \frac{\delta \widetilde{C}_G(u_1, u_2)}{du_1} - u_2 > 0, \qquad h_2(u_1, u_2) = \frac{\delta \widetilde{C}_G(u_1, u_2)}{du_2} - u_1 > 0. \tag{4.24}$$

The presence of tail monotonicity for the whole range of the random variables indicates something stronger than contagion. These properties, called Right Tail Increasing (*RTI*) and Left Tail Decreasing (*LTD*) in Esary and Proschan (1972), imply that $P\{Z_2 > z_2 | Z_1 > z_1\} > P\{Z_2 > z_2\}$ and $P\{Z_2 \leq z_2 | Z_1 \leq z_1\} > P\{Z_2 \leq z_2\}$ respectively, for any pair (z_1, z_2), and therefore are synonymous of contagion and interdependence. Note however that these phenomena do not necessarily appear together. There can exist contagion in the extremes between two random variables without being interdependent, and on the other hand, two interdependent random variables can show weaker links (though stronger than being independent) in distress periods than in calm periods.

The concepts of contagion and interdependence introduced so far regard the intensity of the dependence, that is, the strength of the links between the variables as these go further into the tails. However, other forms of contagion regarding the direction of the dependence are found, in this setup the conditional probability of (4.19) is interpreted as a causality relationship. Contagion in this context occurs when one variable is influencing the other, that is, a large value in one variable is raising the likelihood of a large value in the other variable. Then the relation between the variables must be asymmetric, otherwise there is only an increase in the intensity of the dependence (contagion as defined in 4.3.2). Note however that a condition of the type $P\{Z_2 > z_2 | Z_1 > z_1\} > P\{Z_2 > z_2\}$ is equivalent to (4.20). Moreover, the only difference of the former with a condition as $P\{Z_1 > z_1 | Z_2 > z_2\} > P\{Z_1 > z_1\}$ is given by the marginal distributions. In the case of $H^*(z_1, z_2)$ where the margins are identical Fréchet, both conditional probabilities are identical.

Let us focus instead in the following conditions for contagion spill-over,

$$P\{Z_2 > z_2 | Z_1 > z_1\} \geq P\{Z_1 > z_2 | Z_2 > z_1\}, \tag{4.25}$$

for the upper tails, with $z_2 \geq z_1$, and

$$P\{Z_2 \leq z_2 | Z_1 \leq z_1\} \geq P\{Z_1 \leq z_2 | Z_2 \leq z_1\}, \tag{4.26}$$

for the lower tails, where $z_2 \leq z_1$. These conditions boil down to see if $\widetilde{C}_G(u_1, u_2; \Theta) > \widetilde{C}_G(u_2, u_1; \Theta)$. Consider z_1 a threshold value that determines the extreme events, hence this inequality implies that the likelihood of Z_2 being extreme given that Z_1 is extreme is larger than the likelihood of Z_1 being extreme with Z_2 extreme. In other words, Z_1 is causing Z_2 reaches extreme values. The particular case of equality in the latter expressions represents symmetry of the variables and economically concerns directional interdependence (both economies are affected by the external factors in the same way).

To formalize directional contagion define $gd_v(u) = \widetilde{C}_G(u,v) - \widetilde{C}_G(v,u)$ and introduce the following definition,

Definition 4.3.3. *Suppose Z_1, Z_2 with common Fréchet distribution Ψ_1 and consider z a threshold that determines the extremes of both random variables. Then, Z_1 is influencing Z_2 in the extreme values (contagion effect) if $gd_v(u)$ is strictly positive for all $v > u$ for the upper tail, and for all $v < u$ for the lower tail, with $u = \Psi_1(z)$.*

This definition is analog for Z_2 influencing Z_1 but reversing the signs of the inequality. In terms of the parameters of the copula $\widetilde{C}_G$, there is contagion from Z_1 towards Z_2 if $\eta > 1$, and from Z_2 towards Z_1 if $\eta < 1$.

The definition may be strengthened by imposing a monotonicity condition on $gd_v(u)$. The intensity of this type of contagion can be measured by means of this monotonicity condition. In particular,

Definition 4.3.4. *Suppose the conditions of definition 4.3.3, and Z_1, Z_2 such that there is positive contagion from Z_1 to Z_2. Then, Z_1 is strongly influencing Z_2 in the extreme values (strong contagion effect) if $gd_v(u)$ is an increasing function in v for all $v > u$.*

A characterization of this definition is

$$sc(u,v) = \frac{\delta \widetilde{C}_G(u,v;\Theta)}{dv} - \frac{\delta \widetilde{C}_G(v,u;\Theta)}{dv} > 0. \tag{4.27}$$

The economic interpretation behind lies on irrational increases in the probability that Z_2 becomes extreme given that Z_1 has reached extreme observations (remind that the variables represent innovations).

4.3.1 Estimation of the Copula: Testing Contagion

In general, to estimate the set of dependence parameters Θ of any multivariate distribution function two strategies may be employed. If the marginal distributions are known or can be estimated by valid parametric models the likelihood function for the data is easily derived. If the multivariate distribution function is $H(x_1, x_2) = C(F_1(x_1), F_2(x_2); \Theta)$ the likelihood function is

$$\pounds(\Theta; x_1, x_2) = \sum_{i=1}^{n} log\ f_1(x_{i,1}) + \sum_{i=1}^{n} log\ f_2(x_{i,2}) + \sum_{i=1}^{n} log\ c(F_1(x_{i,1}), F_2(x_{i,2}); \Theta), \quad (4.28)$$

with f_i the marginal density function of F_i, and $c(F_1(x_1), F_2(x_2); \Theta)$ the bivariate density of the copula. The resulting estimates of the dependence parameters are margin-dependent, as well as the parameters of the corresponding marginal distributions. On the other hand the estimates of Θ are free of these effects for nonparametric estimates of the margins. Genest, Ghoudi, and Rivest (1995) show that the estimates derived from a pseudo-likelihood estimation are consistent and asymptotically normal. This method is implemented in two steps. First, the estimates of the marginal distributions are estimated by the respective nonparametric empirical distribution functions. In this way u_i is obtained as $u_i = \hat{F}_{i,n}(x)$, with $\hat{F}_{i,n}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}$, and the log-likelihood for C is

$$\pounds(\Theta; u_1, u_2) = \sum_{i=1}^{n} log\ c(u_{i,1}, u_{i,2}; \Theta). \quad (4.29)$$

In our case, $H^*(z_1, z_2) = \widetilde{C}_G(\Psi_1(z_1), \Psi_1(z_2); \Theta)$, though the marginal distribution functions are known, standard Fréchet, the underlying marginals F_1, F_2 are not, so it is preferable to consider the nonparametric case. Note that $u_i = \Psi_1(z_i)$ boils down to $u_i = F_{i,n}(x_i)$ by construction of Z_i. The log-likelihood function of H^* is calculated as in (4.29). This function, however, does not take an easy-to-handle expression in logs, and the score function does not adopt a closed form. Instead numerical optimization methods are employed to maximize the likelihood.

An appealing property of the copula $\widetilde{C}_G$ is its nested character. It is immediate to see that $\gamma = 0$ is the standard Gumbel distribution, that represents the class of bivariate extreme value distributions. The case of asymptotic independence for the right tail is given by $\theta = 1$ whilst perfect independence is described by $\theta = 1$ and $\gamma = 0$. Finally, η measures the level of asymmetry (exchangeability) of the variables. As a result, it is straightforward to implement tests for the corresponding hypotheses about dependence by means of likelihood ratio tests. The test statistic is

$$\Lambda_n = 2\, log \frac{\underset{\Theta}{sup} \prod_{i=1}^{n} \widetilde{c}_G(u_{i,1}, u_{i,2}; \Theta)}{\prod_{i=1}^{n} \widetilde{c}_G(u_{i,1}, u_{i,2}; \Theta_0)}, \tag{4.30}$$

with Θ_0 the set of parameters under the null hypothesis. The asymptotic distribution of Λ_n is chi squared-distributed with degrees of freedom equal to the difference of the dimensions between Θ and Θ_0.

The nested character of the copula makes immediate testing dependence as well as testing the existence of contagion effects in the data. The corresponding hypotheses tests are $H_0 : \theta = 1,\ \gamma = 0$ vs $H_1 : \theta > 1$ or $\gamma > 0$, and $H_0 : \eta = 1$ vs $H_1 : \eta \neq 1$ for $\gamma > 0$. Meanwhile the existence of intensity contagion and strong directional contagion boil down to study conditions (4.24) and (4.27) respectively plugging the estimated parameters.

4.4 Application: Flight to quality versus Contagion

Financial crises are characterized by dramatic falls in the prices of asset returns for reference markets. The fall in prices of these returns trigger a sequence of negative effects on the prices of the rest of the assets traded in the market by different reasons: a remarkable weight of the asset in the composition of the portfolios, bilateral trade, or a psychological or contagion effect.

It seems logical to think that investors in order to avoid the pernicious effects of the crisis flee towards safe markets: the bonds market. However, sometimes it is not clear the type of market failing and originating the crisis since the overall economic structure collapses. In this situation the refuge in the bonds market may not provide with the desired coverage against losses. The phenomenon of fleeing from the stocks market to the bonds market is known as flight to quality. Measuring this effect is useful in a number of ways: it reflects the links between these markets, in cases of crisis it is useful to identify its sources (financial vs other types), or the causality of the relationship, that is, if bear stock markets imply bull bond markets, or there is some common economic factor producing the co-movements (*e.g.* low interest rates).

In this section this phenomenon is tested for two different pairs of financial indexes: the Dow Jones Corporate 02 Years Bond Index (DJBI02) vs the Dow Jones Industrial Average: Dow 30 Industrial Stock Price Index (DJSI), and the the Dow Jones Corporate 30 Years Bond (DJBI30) Index vs the Dow 30 Industrial Stock Price Index. These series are studied for the period 02/01/1997 − 24/09/2004. The Corporate Bonds Indexes data are taken from the official Dow Jones Index website and the Stock Price Index from *www.freelunch.com*. Sample observations corresponding to public holidays and missing data in either of the series are deleted from both data sets to avoid the incorporation of spurious zero returns and aberrant dependencies, leaving n=1942 observations. The observations considered for the analysis are the logarithmic returns measured in percentage terms and denoted as r_t,

$$r_t = 100\,(logP_t - logP_{t-1}),$$

with P_t the original prices at time t.

The methodology followed in this empirical work starts by filtering the data by univariate models as sketched in (4.1) and analyzing the dependence patterns between the resultant innovation vector $(\varepsilon_1, \varepsilon_2)$ by means of the copula $\widetilde{C}_G$. This copula is sufficient for testing the existence of contagion effects, co-movements, or opposite effects in the tails that are reflected in the set of parameters Θ of the copula $\widetilde{C}_G$.

Tables 4.6.1, 4.6.2 show that *DJBI02* index is well modelled by an AR(1)-GARCH(1,1) model as follows,

$X_{1,t} = 0.00025 + 0.089X_{1,t-1} + \sigma_{1,t}\varepsilon_{1,t}$, with $\varepsilon_{1,t}$ *i.i.d.* $(0,1)$, and
$\sigma^2_{1,t} = 6.194 \cdot 10^{-8} + 0.071\varepsilon^2_{1,t-1} + 0.903\sigma^2_{1,t-1}$.

The *DJSI* Index is modelled by the following pure GARCH(1,1) model (tables 4.6.3, 4.6.4),

$X_{2,t} = \sigma_{2,t}\varepsilon_{2,t}$, with $\varepsilon_{2,t}$ *i.i.d.* $(0,1)$, and $\sigma^2_{2,t} = 3.0012 \cdot 10^{-6} + 0.096\varepsilon^2_{2,t-1} + 0.887\sigma^2_{2,t-1}$.

The bivariate sequence of innovations $(\varepsilon_{1,t}, \varepsilon_{2,t})$ is represented in figure 4.6.1. A first glance to the picture provides some guidance towards the existence of a flight to quality effect between the innovations of *DJSI* and the innovations corresponding to *DJBI02*. The analysis of cross correlation (figure 4.6.2) confirms the existence of opposite shocks in the innovation sequence as well as validates the univariate models proposed to satisfy the assumptions in (4.1).

The copula function $\widetilde{C}_G$ introduced in this chapter is estimated numerically. The parameter estimates for this example are $\hat{\theta}_n = 1.031$, $\hat{\eta}_n = 1$ and $\hat{\gamma}_n = 0.175$. This model fits well the data for different sections of the copula for both margins as can be seen (see figure 4.6.3). The following pictures are derived from $\widetilde{C}_G$ estimated from the data. In this way, figures 4.6.4 and 4.6.5 show negative interdependence between the random variables in the left tail, that becomes stronger in the middle of the bivariate distribution and turns positive in the right tail. Both plots are identical indicating the absence of directional contagion, that is, asymmetric effects between the variables. This is also described in figure 4.6.6. On the other hand it is remarkable the presence of intensity contagion in the left tail (figure 4.6.5). A deeper analysis of figures 4.6.4 and 4.6.5 shows opposite movements in the middle of their domain, that decrease when the variables take larger absolute values. This phenomenon is more pronounced for the extreme negative values that tend to move together, or at least not in opposite directions (contagion without interdependence).

It is convenient not confusing the contagion phenomenon just illustrated that appears when both variables simultaneously take on extreme events in the same tail with the flight to quality. This phenomenon occurs when the extreme values occur in the opposite tails, in particular when *DJBI02* takes positive extreme values and *DJSI* negative extreme values.

Figure 4.6.7 clearly describes the existence of this phenomenon in both tails, that may be interpreted as a substitution effect between these financial sequences when either of the sequences are in crises periods.

The analysis for the pair Dow Jones Corporate 30 Years Bond Index (*DJBI30*) and the Dow Jones Stock Price Index (*DJSI*) yields different results. *DJBI30* is modelled by an AR(1)-GARCH(1,1) model where *DJSI* also enters in the equation. The parameter estimates are displayed in tables 4.6.5 and 4.6.6 and can be summarized as follows,

$X_{1,t} = 0.00037 + 0.063X_{1,t-1} + 0.048X_{2,t-1} + 0.028X_{2,t-2} + \sigma_{1,t}\varepsilon_{1,t}$, with $\varepsilon_{1,t}$ $i.i.d.$ $(0,1)$, and $\sigma^2_{1,t} = 1.375 \cdot 10^{-6} + 0.056\varepsilon^2_{1,t-1} + 0.905\sigma^2_{1,t-1}$.

The pair $(\varepsilon_{1,t}, \varepsilon_{2,t})$ is represented in figure 4.6.8. The cross correlation function (figure 4.6.9) indicates the absence of linear correlation between any lag combination. This graph also assesses the univariate model proposed to describe the dynamics of *DJBI30*. The parameter estimates for $\widetilde{C}_G$ in this case are $\hat{\theta}_n = 1.01$, $\hat{\eta}_n = 1$ and $\hat{\gamma}_n = 0.0003$. This model suits very well to the data (figure 4.6.10). The fitted copula shows that the random variables are weakly interdependent (figure 4.6.11), that is, both innovation sequences, though close to independence, move in the same direction. More formally, the likelihood ratio test introduced in (4.30) calculated for $H_0 : \theta = 1, \gamma = 0$ vs $H_1 : \theta > 1$ or $\gamma > 0$ does not reject the hypothesis of independence. Note from the value of γ the random variables are symmetric and therefore there is not directional contagion. Furthermore, figure (4.6.12) describes absence of intensity contagion in either of the tails reflecting a weakening in the links between the variables in the tails. In the limit these random variables are asymptotically independent in both tails. Finally the flight to quality phenomenon is not present for these two series as shown in figure 4.6.13.

4.5 Conclusions

Contagion and interdependence are different concepts. In this chapter contagion is related to extreme or tail events. Via the theory of copulas, we are able to analyze and test the existence of contagion, its intensity, as well as its causal direction. This is done by creating a new copula, derived from the multivariate extreme theory, that is sufficiently flexible both to describe different patterns of dependence, and model asymmetric effects between markets.

This copula has been applied to study the links between safe and risky markets represented by the Dow Jones Corporate Bond Index (*DJBI*) and the Dow Jones Stock Price Index (*DJSI*). From the point of view of economic fundamentals, the latter index is independent of *DJBI*, while the bonds indexes, *DJBI02* and *DJBI30*, have different behaviors depending on their maturity. The price of *DJBI02* is independent of the evolution of risky markets, actually the conditional mean price is only driven by its own past price, while the conditional variance is well modelled by a GARCH(1,1) model. On the other hand, *DJBI30* is positively influenced by the evolution of *DJSI* reflecting the health of the overall economy.

Regarding the irrational links between the markets reflected in the innovations sequences and modelled by the copula function introduced in this chapter, the conclusions are also different for the corresponding pairs of financial series. The shocks between *DJBI02* and *DJSI* are negatively related. In particular, the flight to quality effect is present indicating a substitution effect between both financial instruments when either of them are through distress periods. It is also remarkable the existence of a contagion effect in the intensity of the dependence in situations of crises in both markets, common negative shocks. On the other hand *DJSI* and *DJBI30* innovation sequences are almost independent. There is no contagion or flight to quality effect.

The conclusion regarding the dependence of these financial series is that while *DJBI02* can serve as refuge for investors fleeing from crises attributed to the stocks markets, *DJBI30* reflects the health of the overall economy, including the stocks markets, and are used by a type of investors not concerned with sharp fluctuations of prices in the stocks markets.

4.6 Appendix: Tables and Figures

Parameter	Value	Standard Error	T Statistic
C	0.00023981	$3.2242e-005$	7.4378
AR(1)	0.075112	0.023976	3.1327
AR(2)	-0.0060864	0.023504	-0.2590
Regress(1)	0.0033575	0.0023594	1.4231
Regress(2)	0.001902	0.0023953	0.7940
Regress(3)	-0.0032545	0.0027541	-1.1817
K	$4.6857e-008$	$1.0825e-008$	4.3286
GARCH(1)	0.92472	0.0088449	104.5485
ARCH(1)	0.055603	0.0053162	10.4591

Table 4.6.1. *Parameter estimates for DJCB02 Index for the period* 02/01/1997−24/09/2004.

Parameter	Value	Standard Error	T Statistic
C	0.000254010	$3.1119e-005$	8.1626
AR(1)	0.089148	0.024137	3.6934
K	$6.1945e-008$	$1.4083e-008$	4.3985
GARCH(1)	0.90361	0.011627	77.7188
ARCH(1)	0.071044	0.0066707	10.6502

Table 4.6.2. *Parameter estimates for DJCB02 Index for the period* 02/01/1997−24/09/2004.

Parameter	Value	Standard Error	T Statistic
C	0.00049681	0.00024775	2.0053
AR(1)	−0.0078747	0.026044	−0.3024
AR(2)	−0.011078	0.023067	−0.4803
Regress(1)	0.0060823	0.0409134	0.1487
Regress(2)	0.035009	0.043498	0.8048
Regress(3)	−0.056755	0.038002	−1.4935
K	$2.9658e - 006$	$8.5241e - 007$	3.4793
GARCH(1)	0.88763	0.011124	79.7956
ARCH(1)	0.095891	0.0092913	10.3205

Table 4.6.3. *Parameter estimates for DJSP Index for the period* 02/01/1997−24/09/2004.

Parameter	Value	Standard Error	T Statistic
C	0.00049118	0.00024454	2.0086
K	$3.0012e - 006$	$8.5464e - 007$	3.5116
GARCH(1)	0.88719	0.010683	83.0474
ARCH(1)	0.096116	0.0084832	11.3302

Table 4.6.4. *Parameter estimates for DJSP Index for the period* 02/01/1997−24/09/2004.

Parameter	Value	Standard Error	T Statistic
C	0.0003741	0.000129	2.9001
AR(1)	0.063049	0.023987	2.6284
AR(2)	−0.0075111	0.023612	−0.3181
Regress(1)	0.04884	0.0097296	5.0197
Regress(2)	0.028364	0.01017	2.7891
Regress(3)	0.0058778	0.011312	0.5196
K	$1.3042e-006$	$4.3895e-007$	2.9712
GARCH(1)	0.90783	0.019353	46.9090
ARCH(1)	0.055566	0.0098548	5.6385

Table 4.6.5. *Parameter estimates for DJCB30 Index for the period* 02/01/1997−24/09/2004.

Parameter	Value	Standard Error	T Statistic
C	0.00037386	0.00012836	2.9127
AR(1)	0.063416	0.023811	2.6633
Regress(1)	0.048314	0.0097464	4.9572
Regress(2)	0.028526	0.010177	2.8031
K	$1.3752e-006$	$4.5783e-007$	3.0038
GARCH(1)	0.90551	0.020033	45.2007
ARCH(1)	0.055866	0.010029	5.5704

Table 4.6.6. *Parameter estimates for DJCB30 Index for the period* 02/01/1997−24/09/2004.

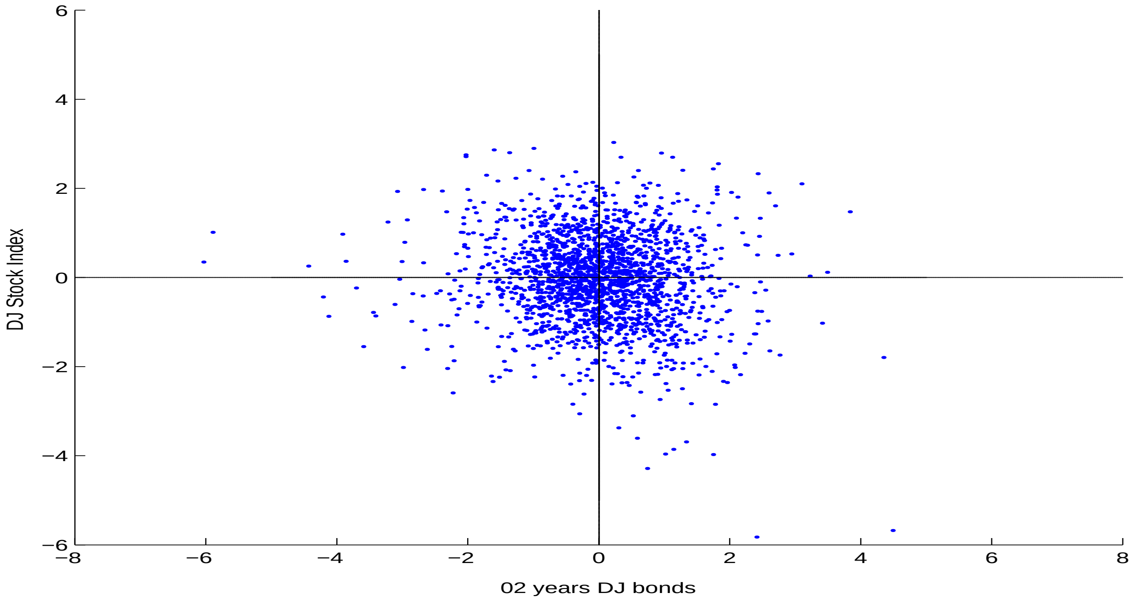


Figure 4.6.1. *Bivariate plot of the innovations sequences of the Dow Jones Corporate 02 Years Bonds and the Dow Jones Stock Index. The observations span the period* 02/01/1997– 24/09/2004, $n = 1942$ *observations.*

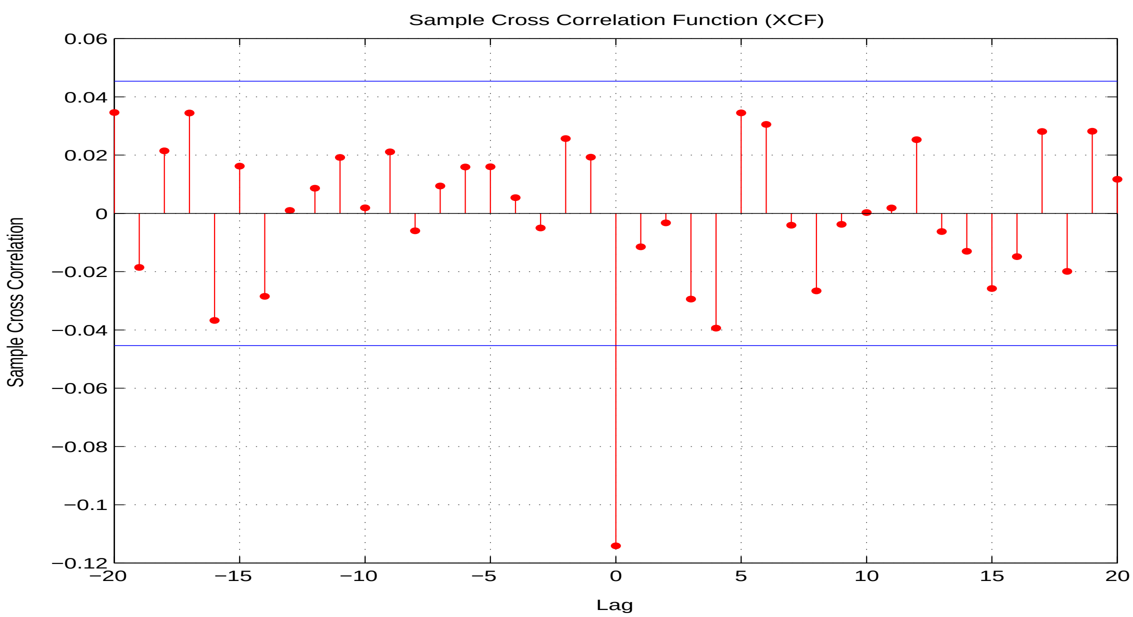


Figure 4.6.2. *Cross correlation for different lags of the bivariate innovation sequence, Dow Jones Corporate 02 Years Bonds and Dow Jones Stock Index, spanning the period* 02/01/1997 – 24/09/2004, $n = 1942$ *observations.*

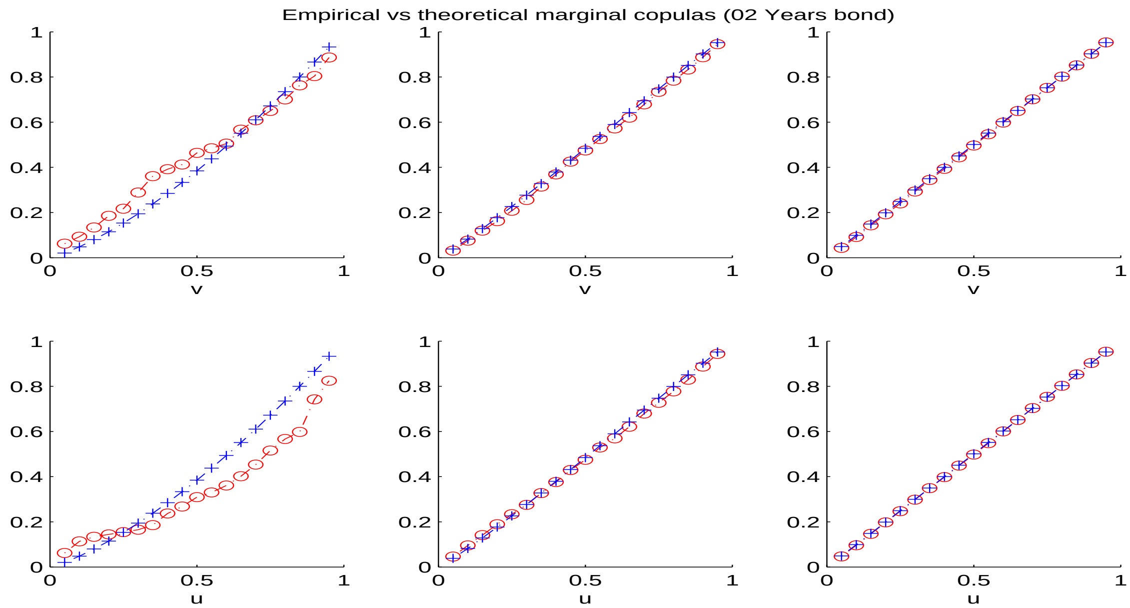


Figure 4.6.3. *Empirical (o−) and theoretical (+−) margins of the cumulative bivariate distribution function. The upper panel describes the vertical sections and the lower panel the horizontal section. The left panels represent* 0.05 *quantile, the middle panels* 0.50 *quantile and the right panels the* 0.0.95 *quantile.*

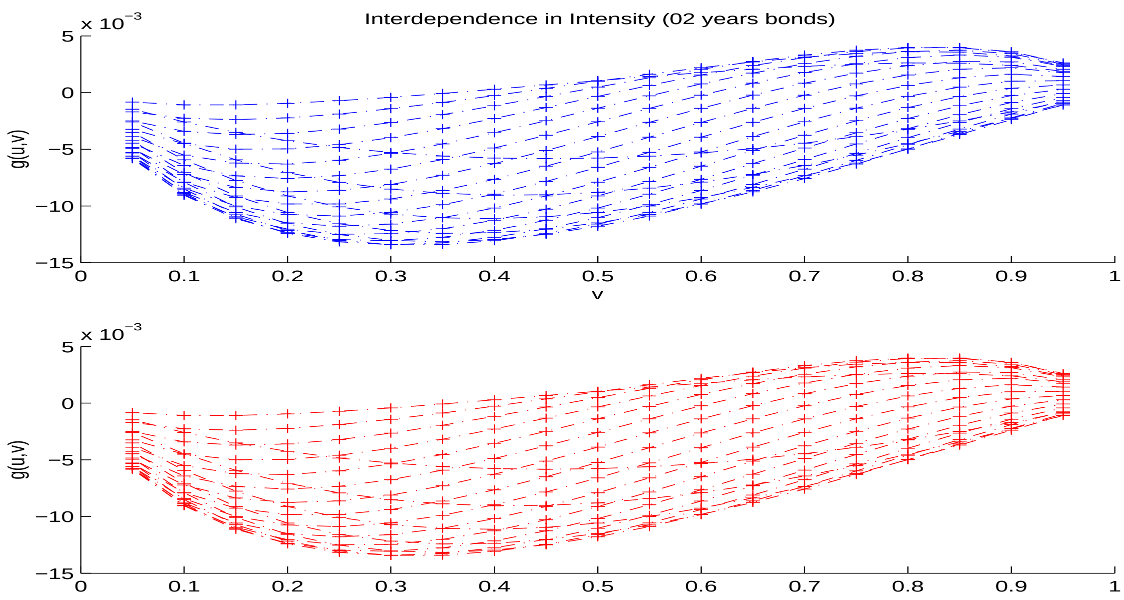


Figure 4.6.4. *The upper panel depicts the function $g(u,v)$ as defined in* (4.23) *plotted against the innovations of DJSI. The lower panel $g(u,v)$ plotted against the innovations of DJBI02.*

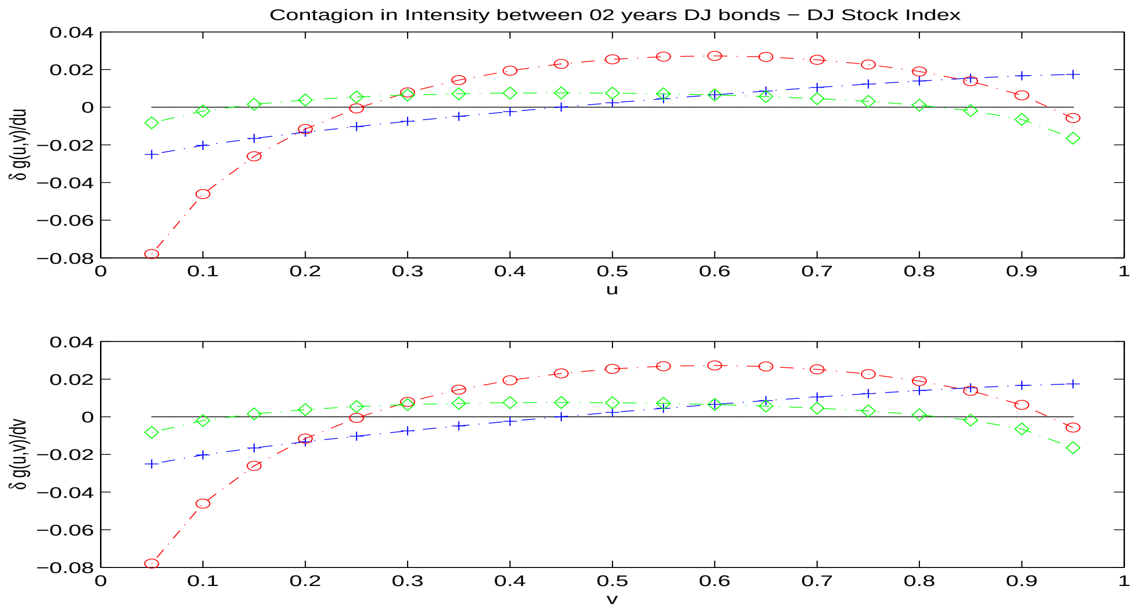


Figure 4.6.5. *The upper panel depicts the function $h_1(u, v)$ as defined in (4.24) plotted against the innovations of DJBI02 and the lower panel depicts $h_2(u, v)$ against the innovations of DJSI.* ($+-$) *represents the 0.05 quantile,* ($o-$) *the 0.50 quantile and* ($\diamond-$) *the 0.95 quantile.*

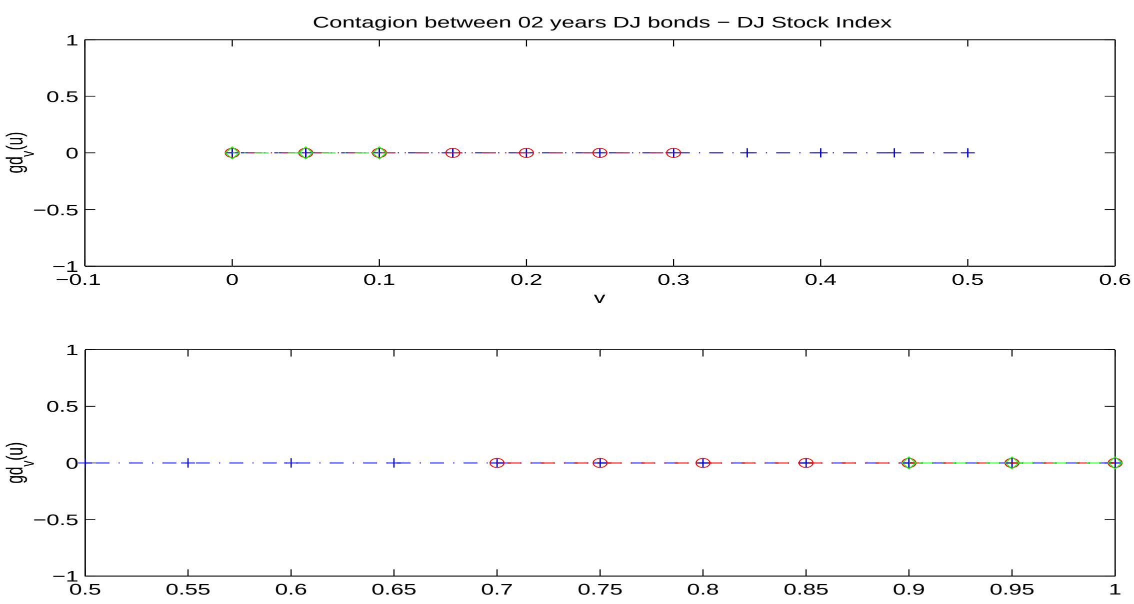


Figure 4.6.6. *The upper panel depicts* $gd_v(u) = \widetilde{C}_G(u, v) - \widetilde{C}_G(v, u)$ *for the lower tail* $(v \leq u)$. ($+-$) *represents* $u = 0.50$, ($o-$) *represents* $u = 0.30$ *and* ($\diamond-$) *for* $u = 0.10$. *The lower panel depicts* $gd_v(u)$ *for the upper tail* $(v > u)$. ($+-$) *represents the* $u = 0.50$, ($o-$) *represents* $u = 0.70$ *and* ($\diamond-$) *for* $u = 0.90$.

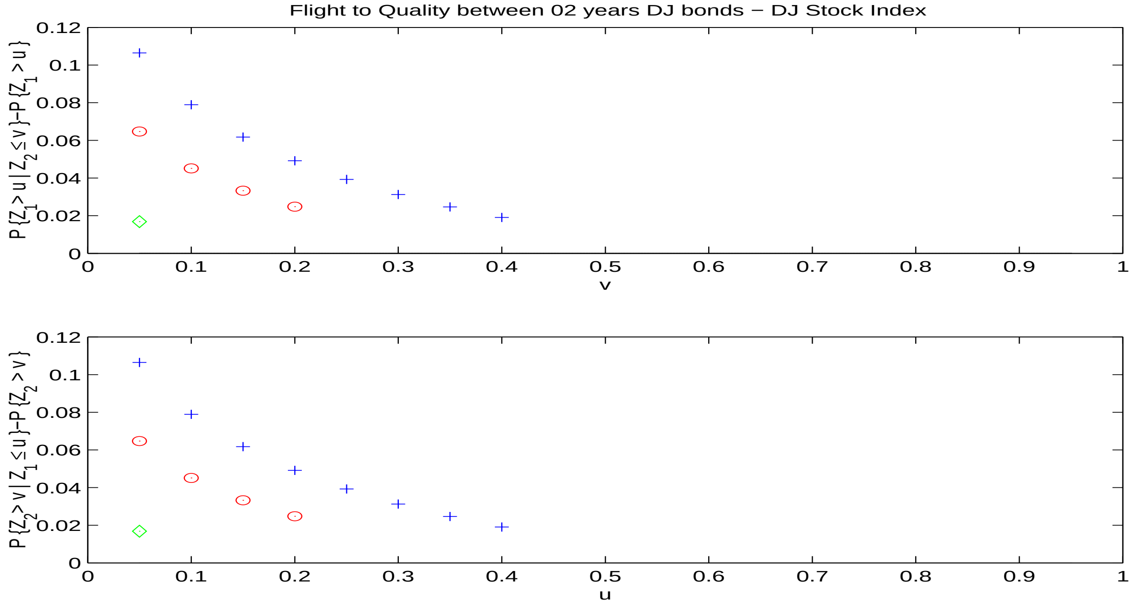


Figure 4.6.7. *The upper panel depicts the flight to quality from DJSI towards DJBI02.* $(+-)$ *represents* $u = 0.60$*,* $(o-)$ *represents* $u = 0.80$ *and* $(\diamond-)$ *for* $u = 0.95$*. The lower panel depicts the flight to quality from DJBI02 towards DJSI.* $(+-)$ *represents* $v = 0.60$*,* $(o-)$ *represents* $v = 0.80$ *and* $(\diamond-)$ *for* $v = 0.95$*.*

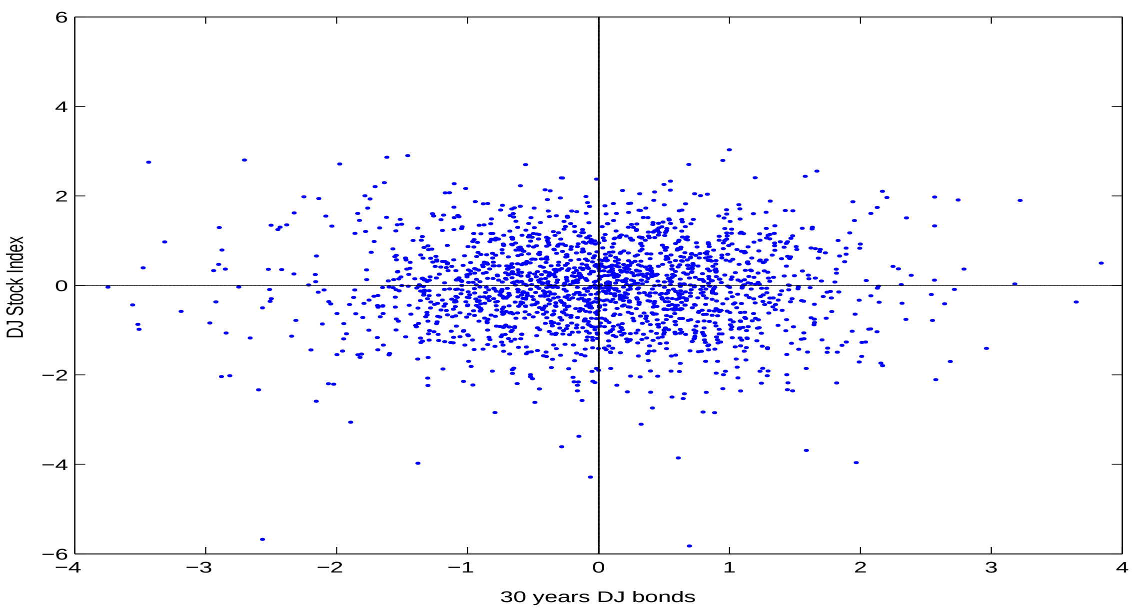


Figure 4.6.8. *Bivariate plot of the innovations sequences of the Dow Jones Corporate 30 Years Bonds and the Dow Jones Stock Index. The observations span the period* 02/01/1997–24/09/2004*,* $n = 1942$ *observations.*

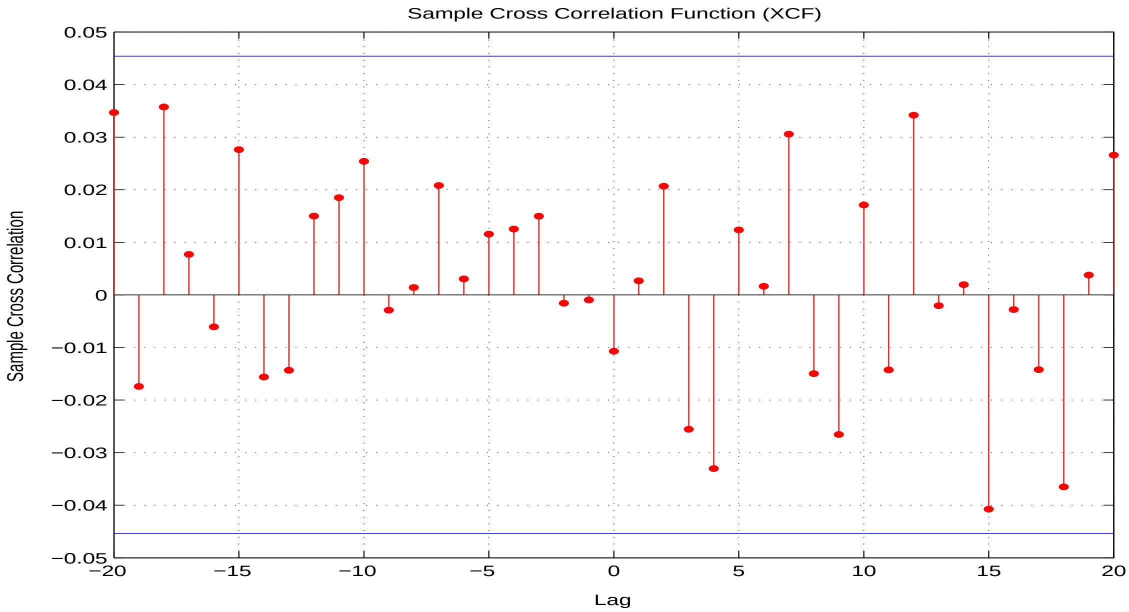


Figure 4.6.9. *Cross correlation for different lags of the bivariate innovation sequence, Dow Jones Corporate 30 Years Bonds and Dow Jones Stock Index, spanning the period* $02/01/1997 - 24/09/2004$, $n = 1942$ *observations.*

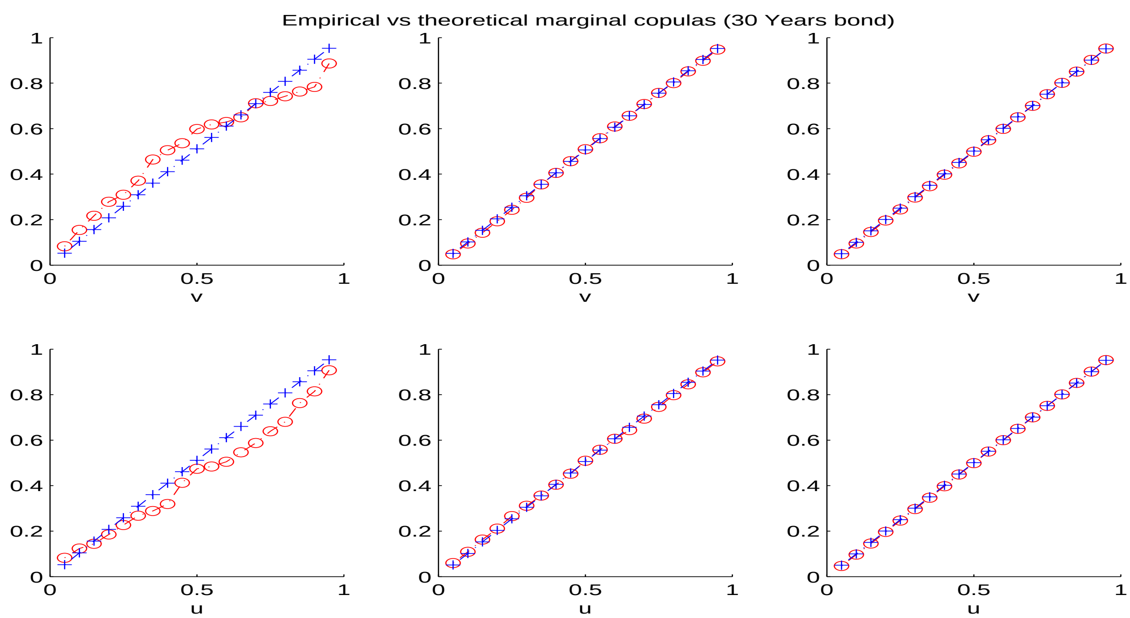


Figure 4.6.10. *Empirical* $(o-)$ *and theoretical* $(+-)$ *margins of the cumulative bivariate distribution function. The upper panel describes the vertical sections and the lower panel the horizontal section. The left panels represent* 0.05 *quantile, the middle panels* 0.50 *quantile and the right panels the* 0.0.95 *quantile.*

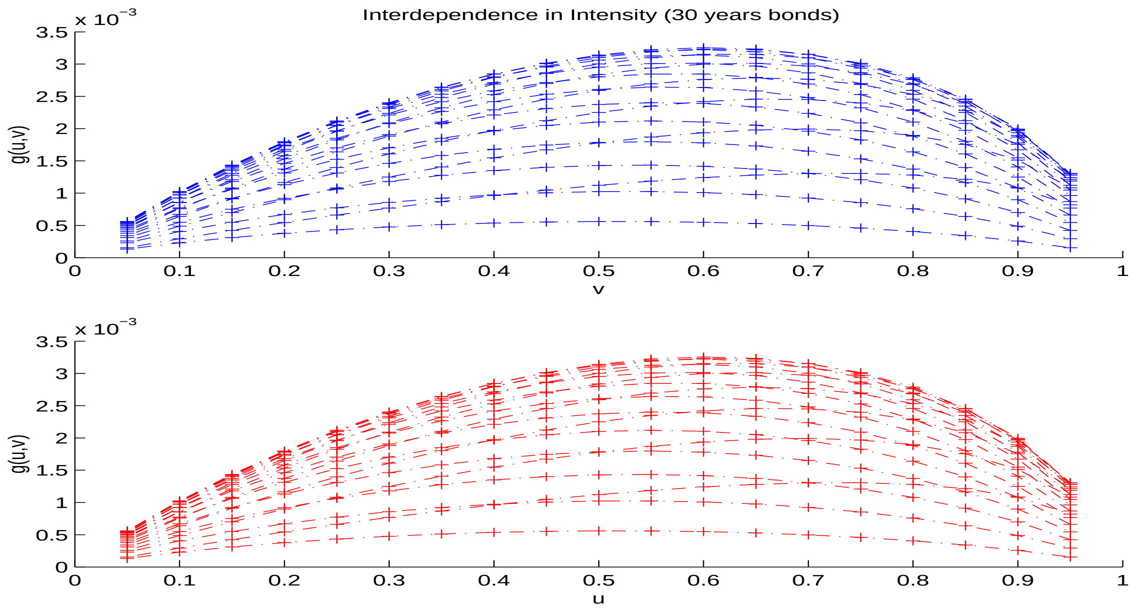


Figure 4.6.11. *The upper panel depicts the function $g(u, v)$ as defined in (4.23) plotted against the innovations of DJSI. The lower panel $g(u, v)$ plotted against the innovations of DJBI30.*

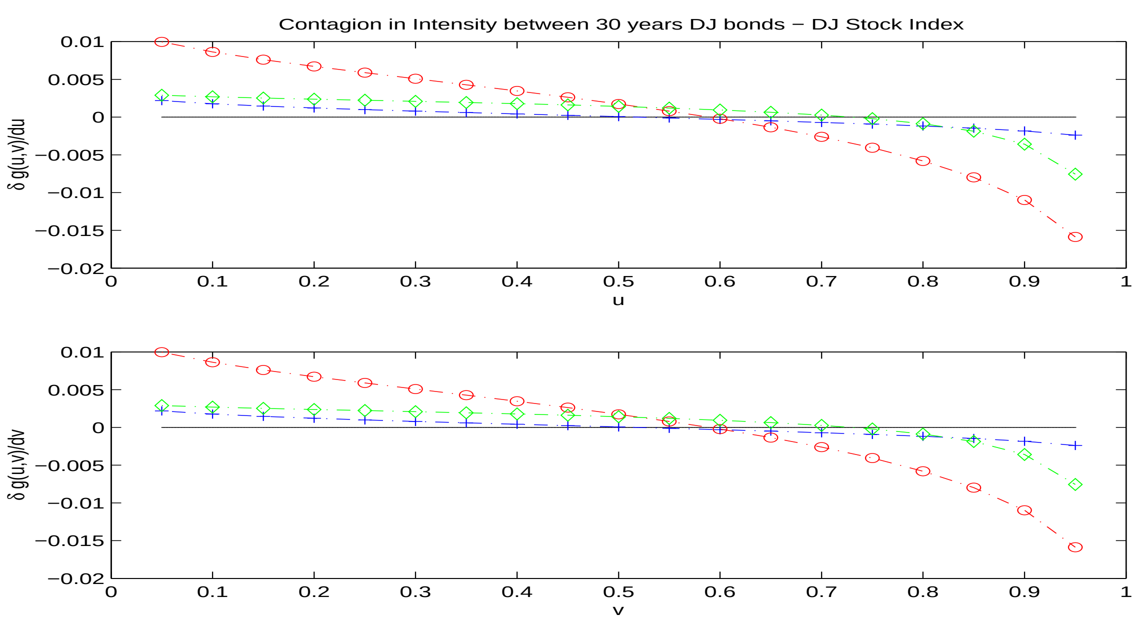
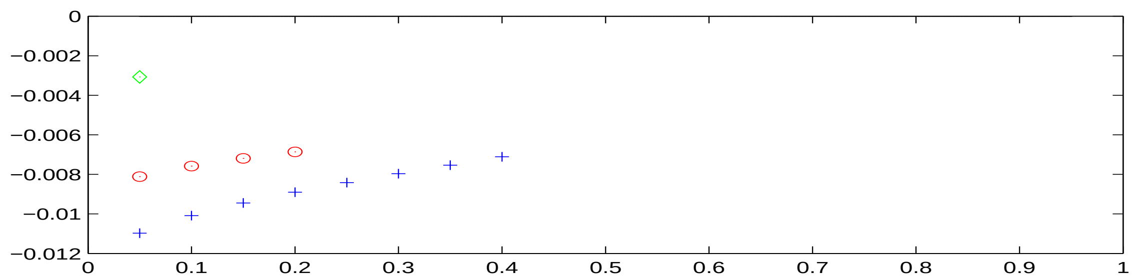


Figure 4.6.12. *The upper panel depicts the function $h_1(u, v)$ as defined in (4.24) plotted against the innovations of DJBI30 and the lower panel depicts $h_2(u, v)$ against the innovations of DJSI. $(+-)$ represents the 0.05 quantile, $(o-)$ the 0.50 quantile and $(\diamond-)$ the 0.95 quantile.*



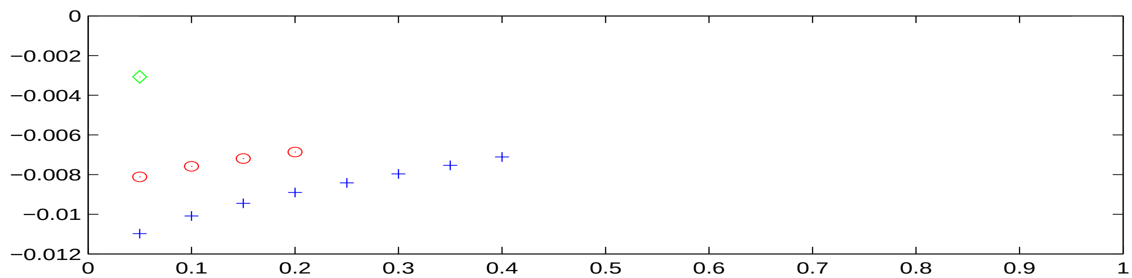


Figure 4.6.13. *The upper panel depicts the flight to quality from DJSI towards DJBI30.* $(+-)$ *represents* $u = 0.60$, $(o-)$ *for* $u = 0.80$ *and* $(\diamond-)$ *for* $u = 0.95$. *The lower panel depicts the flight to quality from DJBI02 towards DJSI.* $(+-)$ *represents* $v = 0.60$, $(o-)$ *for* $v = 0.80$ *and* $(\diamond-)$ *for* $v = 0.95$.

Chapter 5

Future Lines of Research

This chapter sketches possible extensions of the thesis. The discussion is concentrated on three topics. First, the transmission of risk between economies and the effects of contagion in diversification. Second, the transmission of risk in financial sequences involving serial dependence in the extreme observations generated by heavy tails, and finally, the use of EVT techniques for the detection of aberrant observations (outliers) out of the range of the available information and far from being explained by the extreme values of the distribution of the data.

Contagion effects in different aspects of the economy is a promising research area that is accumulating interest from academics as well as from practitioners. This thesis has explored the flight to quality from stocks to bonds markets. Other interesting results involving contagion effects are found in the analysis of banking crises, currency crises or portfolio crises. The immediate aim further this thesis is portfolio risk management. In particular the negative effects of contagion in diversification strategies.

The methodology introduced in Markowitz for portfolio selection (Journal of Finance, 1952, 1991) is based on minimizing the linear combination of the vector of weights of the individual returns and the corresponding variance-covariance matrix. In this way the optimal portfolio is the linear combination with minimum variance. This statistical measure, however, is an adequate tool for measuring risk only under some restrictive conditions: log normal distribution of asset prices (normal distribution of returns), or quadratic utilities. In this world correlation plays the main role to quantify dependence. On the other hand, one of the contributions of this thesis has been highlighting the need of alternative statistical measures for modelling dependence when the distribution of the returns is far from normality. In turn, portfolio decision on the basis of linear measures as variance and correlation may be misleading. The asymmetry observed in financial returns is not reflected in these measures either. The negative skewness implies a higher probability of large negative returns than of the corresponding positive values. In this context, it is important to develop a methodology that is able to account for the risk (negative tail), and for the probability of profits (positive tail) in a separate way.

The extreme value theory techniques are suitable for these problems by their capability to describe the distribution of the returns in the tails. Along with these techniques the bivariate copula function developed in Gonzalo and Olmo (2005) will be the main ingredient to modelling asymmetric relations between the elements of the portfolio. The intention of this research is to move the criterium in portfolio decision from minimizing the variance to minimizing risk, measured as the probability in the negative tail of the portfolio distribution.

The second topic of research regards the transmission of risk in time series. Theories relying on dependence on the second moments have been enormously popular since they arose in the middle eighties. In particular GARCH(1,1) models, due to its simplicity and tractability in practice. Practitioners were very fond of these methodologies as well. As time went by and deficiencies and limitations of this model surged, slight modifications of the base GARCH(1,1) model arose reflecting the stylized facts not described by the original GARCH(1,1). Some of these are the negative skewness of returns or the higher clustering of the negative observations. Different models as EGARCH, IGARCH, etc. were designed to describe these features of financial data. All these methodologies assume conditional gaussian returns where the volatility is time varying and stationary (under some more assumptions) reflecting the dynamics found in financial data.

In the last years however, the common belief that observations are normally distributed has been encountering more and more enemies. The first signs of the disappointment in these models was the need of assuming t-Student distributed observations rather than normally distributed. This family of distributions does not describe very well the empirical features of financial data either. Other authors, however, have raised the use of heavier distributions, for example α-stable distributions with infinite variance. The intention of my future research is something between the gaussian and the heavy tails theories. Berman's condition states that

$$Cov(\varepsilon_t, \varepsilon_{t-j}) \ log\, j \rightarrow 0 \Rightarrow \theta = 1,$$

as $j \rightarrow \infty$, with ε_t the standardized observations. The statistical rejection of the null hypothesis $\theta = 1$ versus $\theta < 1$ implies that either the covariance is not $o(1/log\, j)$ or ε_t is not normally distributed. Provided that the correlation of the financial returns is zero there is no doubt that rejecting $\theta = 1$ amounts to rejecting the gaussian assumption.

Under this evidence it seems appropriate to consider mixtures of distributions as the generating process. The normal distribution for calm periods where the observations are *iid*, and heavy tailed distributions for distress periods. The dependence would be caused by the largest observations that would generate a disturbance effect (panic or euphoria) for a period of time (clusters). No doubt that this theory of transmission of risk in financial sequences is in its infancy and much more attention and theoretical grounds are necessary to construct a reliable theory, but the first steps are built with the hypothesis test for the extremal index and Berman's condition in Olmo (2005a).

In the same way that processes generating clusters of significant large observations may be misspecified and modelled by conditional heteroskedastic volatility models, the opposite result is also remarkable. Isolated outliers may hide true conditional heteroskedasticity and be confused with *iid* processes. This problem, the detection of outliers in financial sequences, is also of interest in risk management. A few large observations may bias the tail estimates of the unconditional distribution of the data. For the conditional case, these significant large observations may affect the estimates of the parameters driving the dependence. It is important then to account for these observations developing a methodology capable of detecting outliers in both cases, for *iid* sequences and for true conditional heteroskedastic volatility models.

A rough definition of outlier is an observation not generated by the generating process of data. In consequence, the first challenge is to know the process. Under *iid* assumptions this boils down to know F if the distribution of data has finite support. Otherwise the former definition is misleading and needs of additional assumptions. On the other hand, for risk management purposes it seems more interesting the detection of outliers in time series. In this case the generating process is not only F but the random process generating the dependence. The current literature in outliers detection tests for time series is based on likelihood ratio tests and regression models between the residuals and the true innovations, see Tsay (1988) and Chen and Liu (1993) for an overview of the methods.

These tests require some information about the process including the type and location of the outlier. In addition, the innovations of the generating process must be gaussian. Under these assumptions, the asymptotic distribution of the test statistic for the null hypothesis is standard normal. If the location of the outlier is not known, however, the null asymptotic distribution is highly non standard and must be calculated by simulation.

The aim of this future research is to benefit of the properties of EVT in order to detect outliers. In particular of the asymptotic results of the distribution of the sample maximum. The hypothesis test and detection rules for outlying observations based on EVT are sketched in the following results.

The three extreme value distributions are related. Given a random variable Y with a Fréchet distribution function (Φ_ξ) there exist some transformations of Y following a Gumbel (Λ) distribution, and a Weibull distribution (Ψ_ξ).

$$Y \sim \Phi_\xi \Leftrightarrow ln\, Y^{1/\xi} \sim \Lambda \Leftrightarrow -Y^{-1} \sim \Psi_\xi, \tag{5.1}$$

with ξ the tail index of F.

Suppose $\{x_n\} = (x_1, \ldots, x_n)$ an *iid* random sequence, and denote $\gamma_n = M_n(\{x_n\})$ for its sample maximum. The statistic used for testing the presence of outliers is $a_n^{-1}(\gamma_n - b_n)$. Under the null hypothesis of no outliers this test statistic follows asymptotically an extreme value distribution of Fréchet, Weibull or Gumbel type. Denoting G_ξ for the extreme value distribution the rejection rule for the null of no outliers will be given by

$$P\left\{a_n^{-1}(\gamma_n - b_n) < G_\xi\right\} < \alpha, \tag{5.2}$$

with α the significance level. Roughly speaking, the likelihood that γ_n is generated by F is negligible. This is measured by the probability that the standardized γ_n belongs to the range of the corresponding extreme value distribution.

Note that in this way the information regarding the location and type of the outlier and the form of the distribution F are not necessary. The only difficulties now arise from the estimation of ξ, a_n, and b_n. Nevertheless by means of (5.1) the test statistic $a_n^{-1}(\gamma_n - b_n)$ can be transformed to follow asymptotically the Gumbel (Λ) distribution, that is parameter free. In consequence, the test statistic is a pivot, although, the estimation of the tail index and the sequences a_n, b_n must be still regarded. The observations of the sequence influencing the estimation of these parameters will be denoted influential observations rather than outliers.

For risk management purposes the extension of the test to dependent processes of GARCH type is also considered. More details on the detection of outliers in the *iid* case as well as for time series can be found in Olmo (2005b).

Bibliography

[1] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001). "The Distribution of Realized Exchange Rate Volatility," Journal of American Statistical Association.

[2] Anderson, T.W., and D.A. Darling (1952). "Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes," Annals of Mathematical Statistics 23, 193 − 212.

[3] Baig, T., and I. Golfajn (1998). "Financial markets contagion in the asian crises," IMF Mimeo.

[4] Barndorff-Nielsen, O.E. and N. Shephard (2002). "Econometric analysis of realised volatility and its use in esimating stochastic volatility models," Journal of the Royal Statistical Society, B, 64 (Forthcoming).

[5] Beran, R. (1986). "Simulated Power Functions," Annals of Statistics 14, 151 − 173.

[6] Berman, S.M. (1964). "Limit theorems for the maximum term in stationary sequences," Annals of Mathematical Statistics 35, 502-516.

[7] Black, F. and M. Scholes (1973). "The pricing of options and corporate liabilities," Journal of Political Economy 81, 637-653.

[8] Bollerslev, T. (1990). "Modelling the Coherence in Sort-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," Review of Economics and Statistics, 72, 498-505.

[10] Chen, C. and L. Liu (1993). "Joint Estimation of Model Parameters and Outlier Effects in Time Series," Journal of the American Statistical Association, 88, 421, Theory and Methods.

[10] Chernick, M.R. (1981). "A limit theorem for the maximum of autorregresive processes with uniform marginal distribution," Annals of Probability 9, 145 – 149.

[11] Coles, S., J. Heffernan, and J.A. Tawn (1999). "Dependence measures for Extreme Value Analysis," Extremes 2:4, 339-365.

[12] Corsetti, G., M. Pericoli, and M. Sbracia (2002). "Some contagion, some interdependence: More pitfalls in Tests of Financial Contagion," Working Paper, Bank of Italia.

[13] Davis, R.A., and S.I. Resnick (1984). "Tail Estimates Motivated by Extreme Value Theory," Annals of Statistics 12, 1467 – 1487.

[14] Diebold, F.X., T. Schuerman and J.D. Stroughair (1998). "Pitfalls and opportunities in the use of Extreme Value Theory in Risk Management," Working Paper Series, Wharton School, Univeristy of Pennsylvania.

[15] de Haan, L. (1976). "Sample extremes: an elementary introduction," Statist. Neerlandica, 30, 161 – 172.

[16] de Haan, L., and J. de Ronde (1998). "Sea and Wind: multivariate extremes at work," Extremes, 1, 7-45.

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

[18] Embrechts, P., A. McNeil, and D. Straumann (1999). "Correlation: Pitfalls and Alternatives," Risk, 12 (5), 11-21.

[19] Engle, R. (1999). "Dynamic Conditional Correlation- A Simple Class of Multivariate GARCH Models." Manuscript.

[20] Engle, R.F., and T. Bollerslev (1986). "Modelling the persistence of conditional variances," Econometric Reviews 5, $1-50$ (with discussion).

[21] Engle, R., and K. Kroner (1995). "Multivariate Simulataneous GARCH," Econometric Theory, 11, 122-150.

[22] Essary, J.D., and F. Proschan (1972). "Relationships among some concepts of bivariate dependence," Annals of Mathematical Statistics 43, 651-655.

[23] Ferro, C.A., and J. Segers (2003). "Inference for clusters of extremes," Journal of the Royal Statistical Society B, 65, $545-556$.

[24] Forbes, K, and R. Rigobon (2001). "No contagion, only interdependence: Measuring stock markets comovements," Journal of Finance 57 (5), 2223-2261.

[25] Genest, C., K. Ghoudi, and L.P. Rivest (1995). "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," Biometrika 82, 3, 543-552.

[26] Gnedenko, B.V. (1943). " Sur la distribution limit du terme d'une série aléatorie," Annals of Mathematics 44, $423-453$.

[27] Goldie, C.M., and R.L. Smith (1987). " Slow Variation with Remainder: Theory and Applications," Quart. J. Math. Oxford Ser. (2), 38, $45-71$.

[28] Gonzalo, J., and J. Olmo (2005). "Contagion versus Flight to Quality in Financial Markets," Working paper, Department of Economics, Universidad Carlos III de Madrid.

[30] Granger, C.W.J. (1969). "Investigating Causal Relationships by Econometric Models and Cross-spectral Methods," Econometrica 37, 24-36.

[30] Granger, C.W.J. (2002). " Some Comments on Risk," Journal of Applied Econometrics 17, $447-456$.

[31] Granger, C.W.J., T. Terasvirta, and A. Patton (2002). "Common Factors in Conditional Distributions," SSE/EFI Working Paper Series in Economics and Finance.

[32] Hall, P. (1982). "On some simple estimates of an exponent of regular variation," Journal of Royal Statistical Society B, 44, 37 – 42.

[33] Hartmann, P., S. traetmans, and C. de Vries (2001). "Extreme market linkages in crisis periods," Discussion Paper Series 2916, Center for Economic Policy Research, www.cerp.org.

[34] Hill, B.M. (1975). "A Simple General Approach to Inference about the Tail of a Distribution," Annals of statistics 5, 1163 – 1174.

[35] Hsing, T. (1991). "Estimating the parameters of rare events. Stochastic Processes and Applications," 37, 117 – 139.

[36] Hsing, T. (1993). "Extremal Index Estimation for a Weakly Dependent Stationary Sequence," Annals of Statistics, 21, 2043 – 2071.

[37] Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.

[38] Kallenberg, O. (1976). Random Measures. Akademie Verlag, Berlin and Academic Press, New York.

[39] Knight, F.H. (1921). "Risk, Uncertainty and Profit," Boston: Houghton Miffin, 1921.

[40] Kolmogorov, A.N. (1933). "Sulla Determinazione Empirica di una Legge di Distribuzione," Giornale del Instituto Italiano degli Attuari 4, 83 – 91.

[41] Leadbetter, M. R. (1983). "Extremes and Local Dependence in Stationary Sequences," Z. Wahrscheinlichkeitstheorie verw. Gebiete 65, 291 – 306.

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

[43] Ledford, A.W., and J.A. Tawn (1996). "Statistics for near independence in multivariate extreme values," Biometrika 83, 1, 169-187.

[44] Ledford, A.W., and J.A. Tawn (1997). "Modelling Dependence within joint tail regions," Journal of Royal Statistical Society B, 49, 475-499.

[45] Lehmann, E.L. (1966). "Some Concepts of Dependence," Ann. Math. Statist. 37, 1137-1153.

[46] Longin, F., and B. Solnik (2001). "Extreme correlation of international equity markets," Journal of Finance 56 (2), 649-676.

[47] Loynes, R.M. (1965). "Extreme Values in Uniformly Mixing Stationary Stochastic Processes," Annals of Mathematical Statistics, 36, $993-999$.

[48] Mashal, R., and A. Zeevi (2002). "Beyond Correlation: Extreme Comovements Between Financial Assets," Paper prepared for the Conference in Extremal Events in Finance held in Montreal, October 2002.

[49] Markowitz, H. (1952)."Portfolio Selection," Journal of Finance 7 (1), 77-91.

[50] Markowitz, H. (1991). "Foundations of Portfolio Theory," Journal of Finance 46, 469-477.

[51] Mason, D., and J.H. Schuenemeyer (1983). "A modified Kolmogorov-Smirnov test sensitive to tail alternatives," Annals of Statistics 3, $933-946$.

[52] Nelsen, R.B. (1999). An introduction to copulas. Lecture Notes in Statistics 139, New York: Springer-Verlag.

[53] O'Brien, G.L. (1974). "The maximum term of uniformly mixing stationary sequences," Z. Wahrscheinlichkeitstheorie verw. Gebiete 30, $57-63$.

[54] O'Brien, G.L. (1987). "Extreme Values for Stationary and Markov Sequences," Annals of Probability, 15, $281-291$.

[55] Olmo, J. (2005a). "Testing the Existence of Clustering in the Extreme Values," Submitted.

[56] Olmo, J. (2005b). "Outliers Detection with Extreme Value Theory," Work in progress.

[57] Patton, A.J. (2001). "Modelling Time Varying Exchange Rate Dependence Using the Conditional Copula," Working Paper 2001-09, Department of Economics, University of California.

[58] Pickands III, J. (1975). "Statistical Inference Using Extreme Order Statistics," Annals of Statistics 1, 119 – 131.

[59] Pickands III, J. (1981). "Multivariate extreme value distributions," In Proceedings of the 43rd Session of the International Statistical Institute 2 859-878. ISI, Voorburg, Netherlands.

[60] Pollard, D. (1984). Convergence of Stochastic Processes. NewYork: Springer-Verlag.

[61] Poon, S.H., M. Rockinger, and J.A. Tawn (2003). "Modelling Extreme-Value Dependence in International Stock Markets," Statistica Sinica 13, 929-953.

[62] Press, W. H. (1992). Numerical Recipes in C: the art of scientific computing. Cambridge University Press, Ed. 2.

[63] Resnick, S. I. (1987). Extreme Values, Regular Variation, and Point Processes. New York: Springer-Verlag.

[64] Romano, J. P. (1988). "A Bootstrap Revival of Some Nonparametric Distance Tests," Jounal of the American Statistical Association 83, 698 – 708.

[65] Rosenblatt, M. (1956). "A central limit theorem and a strong mixing condition," Proc. Nat. Acad. Sci. USA, 42, 43 – 47.

[66] Shiryaev, A. N. (2001). Essentials of Stochastic Finance: Facts, Models, Theory. Singapore: World Scientific Publishing.

[67] Sklar, A. (1959). "Fonctions de répartition à n dimensionset leurs marges," Publ. Inst. Statis. Univ. Paris, 8, 229-231.

[68] Smith, R.L. (1985). "Maximum Likelihood Estimation in a Class of Non-Regular Cases," Biometrika 72, $67-90$.

[69] Smith, R.L. (1987). "Estimating Tails of Probability Distributions," Annals of Statistics 15, $1174-1207$.

[70] Smith, R.L., and I. Weissman (1994). "Estimating the Extremal index," Journal of the Royal Statistical Society B, 56, $515-528$.

[71] Tawn, J.A. (1988). "Bivariate Extreme Value Theory: Models and Estimation," Biometrika 75, 397-415.

[72] Tsay, R.S. (1988). "Outliers, Level Shifts, and Variance Changes in Time Series,"Journal of Forecasting, 7, $1-20$.

[73] von Mises, R. (1936). "La Distribution de la Plus Grande de n Valeurs," Selected papers of Richard von Mises 2, $271-294$. American Mathematical Society. Providence.