THESIS FOR THE DEGREE OF LICENTIATE OF PHILOSOPHY

An Extreme Value Perspective on the Changing Dynamics of Financial Returns

Najaat El-Bayati

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
AND GÖTEBORG UNIVERSITY
Göteborg, Sweden 2003

An Extreme Value Perspective on the Changing Dynamics of Financial Returns

Najaat El-Bayati

ISSN 0347-2809/No. 2003:37

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-41296, Göteborg, Sweden
Telephone +46 (0)31-772 1000

Abstract

In this thesis, we discuss a number of different approaches to analysing time series of multivariate financial returns with an emphasis on extreme movements.

In the first case, we use a non-stationary, unconditional and non-parametric approach to model the volatility. The second method consists of modeling the volatility using a stationary, conditional and parametric approach while in the third case, we turn to a more direct extreme value analysis of the returns.

We exemplify the use of these three methodologies through a detailed analysis of the returns of the daily foreign exchange noon buying rates in New York City for the British Pound (GBP) and the German Mark (DEM) over the last 23 years as well as daily closing prices of the Standard & Poors 500 Composite Stock Price Index (S&P500) and the Dow Jones Industrial Average Index (DJI) over the last 26 years.

Keywords: financial returns, non-stationarity, volatility, covariance, non-parametric regression, Nadaraya-Watson kernel estimator, extremal dependence, GARCH, declustering.

Acknowledgements

My sincere thanks to:

Holger Rootzen and Catalin Stărică, my co-supervisors, for all the time and effort they have put in guiding me throughout this work, for their patience in answering my questions and for the valuable comments they have provided. Thank you!

My colleagues and friends, for making every day so much more fun, both at work and outside work.

My mum, my sisters and my in-laws, for their constant support and encouragement.

Arshad, for always being there for me.

"Liten i magen", for sharing all the ups and downs of these last few hectic months with me and for giving a new perspective to my life.

Finally, the Technical Research Council of Sweden for having partly funded this research.

Contents

Chapter 1

Introduction

Estimating dependence between risky asset returns (changes of prices) is a central issue for many finance applications such as hedging, valuation of exotic options written on more than one asset, and risk management in general. An insight into the *extremal* dependence between assets is important in times of financial crises like the 1997 Asian crisis[1]. So far, the risk associated with extreme price movement is yet to be fully understood and measured.

The conventional dependence measure, Pearson correlation, though widely used is appropriate only for linear association. Since it is calculated as an average of deviations from the mean, the weight given to extreme observations is the same as for all of the other values in the sample. Hence differences in the dependence characteristics for extreme realizations may go unnoticed, with possible serious consequences for a financial institution relying on it as a unique measure of dependence.

As an alternative to the traditional approach, it is possible to draw on statistical developments in Extreme Value Theory (EVT). Most of the applications of EVT to finance are univariate, but multivariate applications are also on the increase. Jansen and de Vries [20] show that the crash of 19th October 1987 may not be an isolated event. Loretan and Phillips [24] use EVT to study the existence of moments of financial returns. Longin and Solnik [23] explore the use of Multivariate Extreme

[1] Academics and market participants often point to the perceived occurrence of contagion or use terminology like the "Asian flu". Others highlight joint shocks and macroeconomic fluctuations, triggering simultaneous crises in several markets or countries.

Value methods for stock market returns. Stărică [37] finds a high level of dependence between the extreme movements of most of the currencies in the EU while Diebold, Schuerman and Stroughair [11] sketch a number of pitfalls associated with the application of EVT techniques to financial data. They put emphasis on the role of small samples and the dependency of financial data, especially as caused by stochastic volatility[2].

Volatility of returns has been the subject of much research. While initially financial models assumed a constant volatility (like the random walk models), later research has shown that this is not a valid assumption. The rapid price evolution on the financial markets is characterized by successive periods of high and low volatility. In these instances, approaches based on constant volatility can result in poor performance of the risk management plan.

Various classes of parametric stochastic volatility models have been proposed and improved with time. While the models fit short financial time series quite well, inconsistencies usually arise for longer time series. Due to their parametric nature, misspecification danger is always there. A common feature of these models is that they model the conditional variance while being usually stationary processes with a constant unconditional variance[3].

An alternative to the fully parametric, stationary and conditional methodology is the non-parametric, non-stationary and unconditional approach (see for instance Drees and Starica [12]). In this case, the unconditional variance is allowed to vary thereby reflecting the changing economic factors affecting the evolution of prices with time. The unconditional variance is estimated using standard non-parametric regression estimators applied to the square returns.

Looking at financial time series, we investigate whether it is only the individual variances which vary with time or if the interaction between the time series also changes with time. Our results seem to indicate that for the time series we have studied, the interaction changes. This is in accordance with the evidence of time

[2] *Stochastic volatility* or simply *volatility* refers to the random changes of the variance as a function of time.

[3] One may question whether stationarity is a reasonable assumption though, especially when working with long financial time series.

varying covariances between financial returns, which has been documented in the literature (see for instance Loretan and Phillips [24]).

Accounting for this changing relationship by estimating a time-varying covariance provided a good description of the dynamics and no linear dependence was observed in the resulting residuals. It also enabled us to remove the serial dependence present in the returns and to get a more homoskedastic series that could be modeled by standard Extreme Value methods. The residuals were independent in the extremes and exhibited stationary extremal dependence. Our analysis seems to show that most of the extreme value dependence was caused by a changing covariance.

The remaining chapters are organized as follows:

Chapter 2 describes some of the properties exhibited by financial data and which need to be taken into account when writing down models. In Chapter 3 we discuss some of the most important time series models employed in volatility modelling as well as the non-parametric models that we focus on in this thesis. Chapter 4 briefly describes univariate Extreme Value Theory and the spectral measure, which is used in the measurement of dependence in multivariate extreme value theory.

In Chapters 5 and 6, we apply the non-parametric approach described in Chapter 3 to the analysis of two sets of bivariate data. The first set consists of daily foreign exchange noon buying rates in New York City for the British Pound (GBP) and the German Mark (DEM). Both cover the period January 11, 1979 to January 25, 2002. They were obtained from the Federal Reserve Statistical Release [16] and were all expressed in terms of US dollars. The second set consists of daily closing prices of the Standard & Poors 500 Composite Stock Price Index (S&P500) and the Dow Jones Industrial Average Index (DJI), ranging each from January 03, 1977 to March 06, 2003. Both are stock market indexes for the US and were downloaded from yahoo finance.

In Chapter 7, we use a conditional stationary parametric framework to analyze the two data sets while in Chapter 8, we investigate the extremal dependence directly on the returns. Finally, we conclude in Chapter 9.

Chapter 2

Stylized Features of Financial Data

We begin with presenting a picture of the empirically observed features of financial data, commonly known as "stylized facts".

For notational convenience, we make the following definitions.

If (P_t) denotes the exchange rates or the stock's closing values, the daily logarithmic price differences (log-returns or returns for short) are defined as

$$r_t = ln(\frac{P_t}{P_{t-1}}) = lnP_t - lnP_{t-1}.$$

By Taylor's formula

$$ln\frac{P_t}{P_{t-1}} = ln(1 + \frac{P_t - P_{t-1}}{P_{t-1}}) \approx \frac{P_t - P_{t-1}}{P_{t-1}}.$$

.

In other words, the log-returns are close to the net return series $R_t := \frac{P_t - P_{t-1}}{P_{t-1}}$.

Subsequently, we will concentrate on the study of r_t.

Shares, stock indices or foreign exchange rates all tend to exhibit similar properties after the transformation above. Appropriate models for financial log-returns data have conventionally tried to take into account these properties, commonly known as "stylized facts".

1. **Fat tails**

   Fat tails mean that the distribution of the returns has tails which are regularly varying[1]. Hence they are, in particular, much heavier than the tails of a normal distribution.

   The most celebrated model in mathematical finance is the Black and Scholes (1973) model which assumes normally distributed log returns. The assumption of normally distributed returns has, however, long been a disputed topic. If models for returns do not have enough probability mass in the tails, it can e.g., result in seriously mispriced hedge strategies which do not uncover all of the risk concerned with holding a specific portfolio, since the possibility of very large returns will not be taken into account.

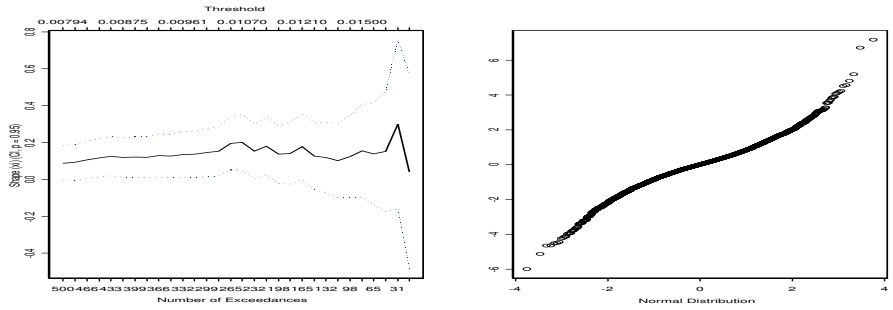

Figure 2.1: Left: *Tail estimation for the GBP log-returns data based on the MLE of the shape parameter* $\xi = 1/\alpha$ *of the GPD (Generalized Pareto Distribution) (see Appendix B), as a function of the number of k upper order statistics (lower horizontal axis) or equivalently of the threshold u (upper horizontal axis). The upper and lower lines constitute asymptotic* 95% *confidence bands.* Right: *QQ-plot of the GBP log-returns data against the normal distribution whose mean and variance were estimated from the returns.*

The left plot of Figure 2.1 with positive values of ξ (see Section 4.1 for its

[1] In other words, there is $\alpha > 0$ such that

$$P[X > x] = x^{-\alpha} L(x), x > 0$$

where $L(x)$ is a slowly varying function, i.e., $L(cx)/L(x) \to 1$ for all $c > 0$ as $x \to \infty$.

implication) and the shape of the QQ-plot, curving down at the left and up at the right, are both an illustration of the heavy tail feature.

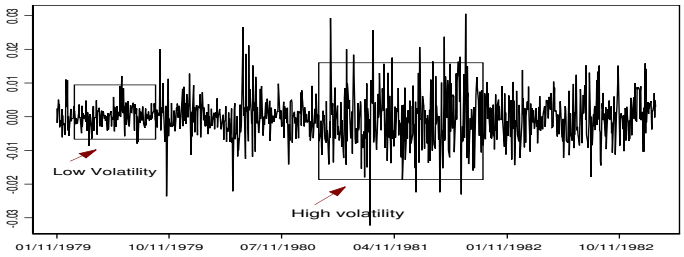

Figure 2.2: *DEM returns from* $01/11/1979$ *to* $10/11/1982$ *where regions of high and low volatility are highlighted.*

2. **Volatility changes with time.**

   This is illustrated by the plot of the returns in Figure 2.2. The regions of high and low volatility are highlighted by the rectangles. It can also be observed that the large price changes come in bulks, or that high threshold exceedances appear in clusters. Proposing statistical models which describe this feature has been the aim of much research activity. For models used in the econometrics literature see the surveys by Bollerslev, Chou, and Kroner [3] or by Taylor [40].

3. **Quasi long range dependence**

   While financial returns hardly show any serial correlation, this does not imply that they are independent especially if we study appropriate functions of the returns. One such function is the square. However, in some cases, like the foreign exchange rates, the 4th moment might not exist (see e.g., Guillaume *et al.* [18]). Then the correlation of the squares is not defined. Another option is to look at the absolute value of the returns. This function compares the returns size-wise and a strong autocorrelation suggests serial dependence.

   We illustrate the above using the log-returns obtained from the GBP and

DEM foreign exchange data sets.

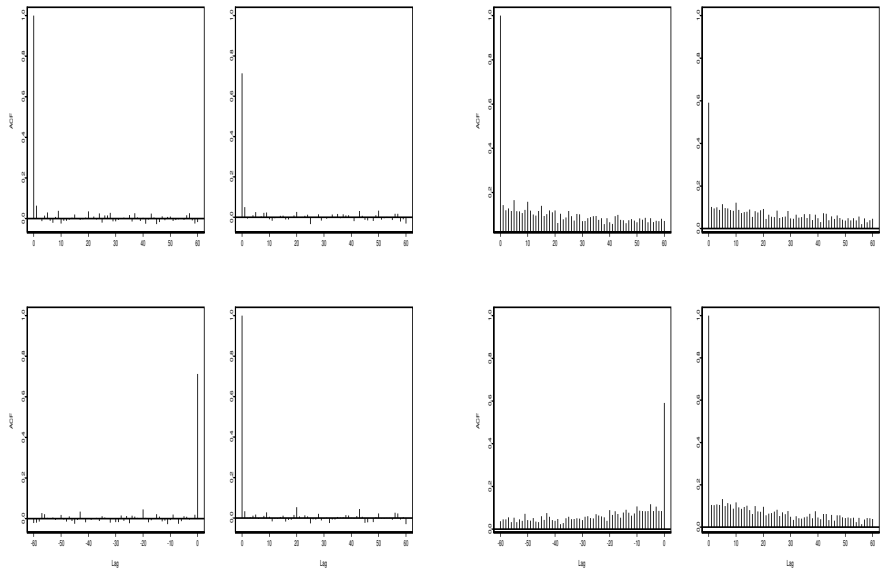

Figure 2.3: Left four plots: *SACFs (Top Left/Bottom Right) and SCCFs (Top Right/Bottom Left) for the GBP and DEM log-returns.* Right four plots: *SACFs (Top Left/Bottom Right) and SCCFs (Top Right/Bottom Left) for the absolute GBP and DEM log-returns.* Note: *No confidence intervals are displayed since the dependency structure in the data is unknown.*

The sample autocorrelation (SACF) $\rho_{n,r}$ of the returns (Top Left/Bottom Right of Left plots in Figure 2.3) is negligible at almost all lags, while the SACF $\rho_{n,|r|}$ of the absolute values $|r_t|$ (Top Left/Bottom Right of Right plots in Figure 2.3) are different from zero for a large number of lags. They instead stay almost constant and positive for large lags, seemingly implying a *long range dependence (LRD) effect*[2] in the volatility.

[2] A stationary process X_t is *long range dependent* if

$$\sum_{h=0}^{\infty} |\rho_X(h)| = \infty \tag{2.1}$$

Note that no confidence intervals are drawn for the correlations since they depend on specifying the type of dependency present in the data. Interpreting the LRD effect is therefore difficult. The behaviour displayed in Figure 2.3 could be evidence of stationary long memory or it could instead be a sign of non-stationarity in the second moment structure (refer to Mikosch and Stărică ([27], [28])).

Having highlighted some of the common features of financial returns, we proceed in the next chapter to introduce some of the models proposed in the literature in order to capture these "stylized facts".

where $\rho_X(h)$ is the sample ACF of X at lag h.

Chapter 3

Volatility Modelling

In this chapter, we turn towards volatility estimation and discuss two approaches. The first is parametric and considers volatility as an unobservable variable to be specified by some stochastic process. The other approach is non-parametric and measures volatility directly from the data, using a transformation of the returns.

Within the first category of models, we look at those for conditional heteroskedasticity and distinguish between the ARCH (autoregressive conditionally heteroskedastic) and SV (stochastic volatility) models.

3.1 Models for Conditional Heteroskedasticity

3.1.1 ARCH Models and Stochastic Volatility Models

The ARCH and the standard Stochastic Volatility (SV) Models share the multiplicative structure

$$r_t = \mu + X_t,$$

where

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{N},$$

with Z_t being an i.i.d. sequence, independent of the volatility process σ_t for fixed t with $E[Z_t] = 0$ and $V[Z_t] = 1$. σ_t and X_t are assumed to be strictly stationary.

For general ARCH (p) models, σ_t is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \quad t \in \mathbb{N},$$

with fixed non-negative constants α_i .

The most well-known discrete time SV model is the log-normal SV model by Taylor [39], which is given by

$$\sigma_t = exp(\nu_t/2),$$

where ν_t is a stochastic process, independent of Z_t and

$$\nu_t = \alpha_0 + \alpha_1 \nu_{t-1} + \eta_t,$$

where the η_t's are independent identically distributed normal random variables.

For the ARCH processes, X_t satisfies

- $E[X_t|Z_{t-1}, Z_{t-2}, \cdots] = 0$. (constant conditional mean).
- $V[X_t|Z_{t-1}, Z_{t-2}, \cdots] = \sigma_t^2$. (time-varying conditional variance).

The corresponding set of properties hold for the SV processes but the conditioning variable is σ_t instead of $Z_{t-1}, Z_{t-2}, \cdots$.

ARCH models are particularly simple to estimate while SV models are easier to analyse from a theoretical point of view. A thorough treatment of SV models can be found in Ghysels, Harvey and Renault [17] or Shephard [35], where the statistical aspects of both ARCH and SV models are discussed.

The ARCH models, developed by Engle [13], were conceived especially for modeling time series that exhibit a changing volatility (for a comprehensive review of the theory of ARCH processes and their applications to financial data, see for instance Bollerslev *et al.* [3]). The considerable interest shown in them is due to their descriptive successes and the availability of maximum likelihood estimates.

The ARCH models have excess kurtosis (Taylor [39]). Furthermore, they show some persistence in the squared autocorrelations depending on the size of p. In

empirical applications of the ARCH (p) model large p's are usually needed. In an attempt to overcome this problem, the Generalized ARCH model, GARCH ($p, q \in \mathbb{N}$), was introduced independently by Bollerslev [2] and Taylor [39] and has become widely used in practice. It shares the same multiplicative structure as the ARCH model above but with σ_t obeying the following relation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{N}, \tag{3.1}$$

with fixed non-negative constants α_i and β_j and where σ_t and X_t are assumed to be strictly stationary.

On the other hand, ARCH and GARCH do not model what is known as *leverage*, which is the fact that volatility is negatively correlated with changes in stock returns in the sense that a dip in stock returns tends to give rise to volatility and vice versa. To address this feature, Nelson [31] introduced the Exponential GARCH (EGARCH) model

$$log\ \sigma_t^2 = \alpha_t + \sum_{i=1}^{\infty} \beta_i\ g(X_{t-i}), \quad \beta_1 \equiv 1,$$

where

$$g(X_t) = \theta_{X_t} + \gamma(|X_t| - E|X_t|).$$

The above-mentioned models are only a few of the ARCH type models which have been considered in the econometrics literature (see Bollerslev, Engle and Nelson [4] for more).

In many financial applications, the GARCH (1,1) model, the simplest GARCH (p,q) model, is sufficient to capture the variance dynamics in the data. This can be determined by looking at the structure of the fitted residuals. The latter often behave very much like an i.i.d. sample, as is expected from the model (see Mikosch [25]).

However, the performance of the GARCH (1,1) model is not free of controversy. For example, there is less dependence in the tails for the returns than for the GARCH (1,1) model (refer to Mikosch [25]). If the data was following a GARCH (1,1) model, we would expect the ACF of the absolute values to decay exponentially to 0 (refer to remark 6.2. in Mikosch [25]). But this is not the case, as seen by the almost constant sample ACFs for the absolute returns in Figure 2.3.

Two other controversial issues are the *long range dependence (LRD) effect* (see (2.1)) and the *IGARCH effect* present in medium and large sample sizes of returns.

The IGARCH effect refers to the case where the sum of the estimated parameters of a GARCH model (estimates for α_i and β_j from equation (3.1)) adds up to a number close to one. The empirical evidence for the IGARCH effect comes from considering long return series (Bollerslev *et al.* [3]). It implies an infinite variance of the log-returns. This is in contradiction though to the finite variance result which one typically observes from the tail analysis on the returns (see de Haan *et al.* [8] and Chapter 6 in Embrechts *et al.* [15]).

While it has been suggested that both the long range dependence (LRD) and the IGARCH effects could be caused by non-stationarity of the returns data (see for instance Mikosch and Stărică ([26], [27])), no agreement has been reached in the literature on the explanation for these effects.

All the volatility models discussed so far assume that the unconditional distribution of asset returns is constant over time. An alternative methodology is based on the assumption of non-stationarity of the distribution of returns. It has however a long history and can be traced back to Hsu, Miller and Wichern [19]. They characterized asset returns as non-stationary processes with discrete shifts in the unconditional variance. This modelling approach is supported by Longin and Solnik [22] who provide some evidence that the unconditional correlation between markets increases during highly volatile periods. Also, Loretan and Phillips [24] reject the hypothesis that stock returns are covariance stationary.

In the non-parametric, non-stationary and unconditional modeling framework,

the univariate model of interest is

$$r_t = \mu + \sigma(t)\,\varepsilon_t, \;\; t = 1, 2, \cdots, n,$$

$$\varepsilon_t \;\text{ iid}, \;\; E(\varepsilon_t) = 0, \;\; \mathit{Var}(\varepsilon_t) = 1,$$

$$\sigma_t, \text{ t = } 1, 2, \cdots, n, \text{ a smooth, deterministic function,}$$

with the volatility being estimated using a two-sided evaluation weighted (Nadaraya-Watson) estimator as

$$\widehat{\sigma}^2(t) \;\; := \;\; \frac{n^{-1}\sum_{i=1}^{n} K_h(i-t)(r_i - \widehat{\mu})^2}{n^{-1}\sum_{i=1}^{n} K_h(i-t)},$$

where $K_h(.) \;\; = \;\; h^{-1}K(./h), h$ is the bandwidth, K is the normal kernel (see Drees and Stărică [12]).

We apply a similar non-parametric regression-type modelling approach to two bivariate sets of returns. In the first case, we model the changing variance as a deterministic, smooth function of time.

3.2 Regression Model I

Denote by $\mathbf{r}_t$ the *m*-dimensional column vector of returns at time *t*. The model assumes that the returns satisfy

$$\mathbf{r}_t = \mu + \mathbf{S}(t)\varepsilon_t, \quad t = 1, 2, \cdots, n \quad \text{where } \mu \text{ is a constant mean,}$$

ε_t is an i.i.d. sequence of vectors with possibly dependent marginals,

$$\mathbf{S}(t), \quad t = 1, 2, \cdots, n, \quad \text{is an invertible } \textit{diagonal} \text{ matrix.}$$

$\mathbf{S}(t)$ is thought of as a smooth, deterministic function of time. The diagonal entries of $\mathbf{S}^2(t)$, corresponding to the volatilities, are estimated by standard non-parametric regression methods for non-random, equidistant design points *t* using each of the square return series. We use the two-sided evaluation weighted (Nadaraya-Watson) estimator given by

$$\widehat{\sigma}^2(t) \;:=\; \frac{n^{-1}\sum_{i=1}^{n} K_h(i-t)(r_i - \widehat{\mu})^2}{n^{-1}\sum_{i=1}^{n} K_h(i-t)}, \tag{3.2}$$

where $K_h(.) = h^{-1}K(./h)$, h is the bandwidth and K is an exponential kernel defined by

$$K(u) = \frac{1}{c} exp^{-5|u|} \quad , |u| \leq 1, \tag{3.3}$$

where c is chosen so that K(u) is a density function. The value 5 in the kernel comes from prior experimentation with exponential filters. The estimates for the volatility are obtained at $t = h+1, \cdots, n-h$.

The residuals are obtained as

$$\widehat{\varepsilon_t} = \widehat{\mathbf{S}}^{-1}(t)(\mathbf{r}_t - \widehat{\mu}), \quad t = b+1, \cdots, n-b,$$

where $\widehat{\mathbf{S}}(t)$ is a square root of the estimate of the diagonal matrix $\mathbf{S}^2(t)$. It is obtained by taking the square root of the diagonal entries of $\mathbf{S}^2(t)$.

$\widehat{\mu} := n^{-1}\sum_{t=1}^{n} \mathbf{r}_t$ is the estimated mean of the returns.

3.3 Regression Model II

In the second set-up,

$$\mathbf{r}_t = \mu + \mathbf{S}(t)\epsilon_t, \quad t = 1, 2, \cdots, n \text{ where } \mu \text{ is a constant mean,}$$

$$\epsilon_t \text{ is an i.i.d. sequence of vectors with independent marginals,}$$

$$\mathbf{S}(t), \quad t = 1, 2, \cdots, n, \text{ is an invertible matrix.}$$

Note: The difference to Regression Model I is that

1. S(t) is no longer diagonal.
2. ϵ_t has independent components.

$\mathbf{S}(t)$ is again thought of as a smooth, deterministic function of time. $\mathbf{S}^2(t)$, which corresponds to the covariance matrix, is estimated by standard non-parametric regression methods for non-random, equidistant design points *t* using the series $(\mathbf{r}_t - \widehat{\mu})(\mathbf{r}_t - \widehat{\mu})', \ t = 1, \cdots, n$ so that we estimate not only the volatilities but also the covariance between the two sets of returns. We use the two-sided evaluation weighted (Nadaraya-Watson) estimator of $\mathbf{S}^2(t)$ given by

$$\widehat{\mathbf{S}}^2(t) \ := \ \frac{n^{-1}\sum_{i=1}^{n} K_h(t-i)(\mathbf{r}_{t-i} - \widehat{\mu})(\mathbf{r}_{t-i} - \widehat{\mu})'}{n^{-1}\sum_{i=1}^{n} K_h(t-i)}, \quad t = 1, 2, \cdots, n, \ (3.4)$$

where $K_h(.) \ = \ h^{-1}K(./h), h$ is the bandwidth and K the same exponential kernel as in (3.3).

The residuals are defined as

$$\widehat{\epsilon_t} = \widehat{\mathbf{S}}^{-1}(t)(\mathbf{r}_t - \widehat{\mu}), \quad t = b+1, \cdots, n-b,$$

where $\widehat{\mathbf{S}}(t)$, the square root of the estimate of $\mathbf{S}^2(t)$, is calculated using the eigenvalue decomposition method. This symmetric square root is used since no hierarchy is present in our bivariate data sets. Otherwise, in the presence of hierarchy, one can use the Cholesky decomposition instead and work with an upper triangular matrix as square root.

Chapter 4

Extreme Value Theory and Extremal Dependence

This chapter deals with Extreme Value Theory and discusses the estimation of the spectral measure. The latter is used in the investigation of the extremal dependence between the coordinates of the residuals.

The numerous approaches by which extreme values may be statistically modelled separate into two forms: methods for maxima over fixed intervals and methods for exceedances over high thresholds. We outline the fundamental aspects of the first method here. Further details about the theory covered in this chapter can be found in Embrechts, Klüppelberg and Mikosch [15] and Resnick [34], Chapter 5.

4.1 The Block Maxima Method

The limit theory for the maximum of a sample of n independent and identically distributed random variables is based on a location-scale normalization of the maximum so that its distribution is non-degenerate as $n \to \infty$.

As an illustration, let (X_i) be a sequence of iid random variables with common distribution F, and denote their partial maxima by $M_n = max(X_1, \cdots, X_n) = \bigvee_{i=1}^{n} X_i,\ n \geqslant 1$.

The Fisher-Tippett Theorem (see Embrechts, Klüppelberg and Mikoch [15]) then states that there exists only three location-scale families of *Extreme Value Distributions G*, namely the Frechet, Weibull or Gumbel distributions, for which

one can find constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} P(c_n^{-1}(M_n - d_n) \leqslant x) = G(x),\ x \in \mathbb{R},$$

and the distribution F is then said to be in the maximum domain of attraction of G ($F \in MDA(G)$).

In other words, provided a non-degenerate limit can be achieved, it follows that, whatever the distribution of the original variables, the limiting distribution of the maximum belongs only to this small set of possible distributions.

These distributions can all be written as the *Generalized Extreme Value Distribution* which is given by

$$G_\xi(x;\mu,\sigma,\xi) = \begin{cases} exp(-(1+\xi(x-\mu)/\sigma)^{-1/\xi}), & \text{if } \xi \neq 0, \\ exp(-exp(-(x-\mu)/\sigma)), & \text{if } \xi = 0 \end{cases}$$

(in the range $1 + \xi(x-\mu)/\sigma > 0$ if $\xi \neq 0$). Here $\sigma > 0$ and μ, ξ may be any real numbers.

The parameter ξ is called the shape parameter and may be used to model a wide range of tail behaviour.

The Generalized Extreme Value Distribution with $\xi < 0$ is called the Weibull distribution Ψ_α. An example of an $MDA(\Psi_\alpha)$ is the uniform distribution on $(0,1)$. The case $\xi = 0$ is referred to as the Gumbel distribution. The normal distribution belongs to its maximum domain of attraction. The third family, where $\xi > 0$, is defined by the Frechet Extreme Value Distribution Φ_α with a positive *tail index* α, where $\alpha = 1/\xi$. Its distribution function is given by $\Phi_\alpha(x) = exp(-x^{-\alpha})$ for $x > 0$.

A characterisation of the maximum domain of attraction of Φ_α is that $F \in MDA(\Phi_\alpha)$ if and only if the right tail $\overline{F} = 1 - F$ is a regularly varying function[1] with index $-\alpha$. Examples of such distributions are the stable distributions and the Student-t distributions.

[1] Recall that in one dimension a random variable X or its distribution is said to be regularly varying if there is $\alpha > 0$ such that

$$P[X > x] = x^{-\alpha}L(x), x > 0$$

where $L(x)$ is a slowly varying function, i.e., $L(cx)/L(x) \to 1$ for all $c > 0$ as $x \to \infty$.

We now move to the multivariate context using the main convention that operations on vectors and relations between vectors are interpreted componentwise.

If $\mathbf{X}_n = (X_{n1}, \cdots, X_{nd}), n \geq 1$ are $\mathbb{R}^d$-valued random variables, define the maximum to be

$$\bigvee_{i=1}^{n} \mathbf{X_i} := (\bigvee_{i=1}^{n} X_{i1}, \cdots, \bigvee_{i=1}^{n} X_{id}),$$

the vector of componentwise maxima.

A multivariate distribution $G(\mathbf{x}) = G(x_1, \cdots, x_d)$ is then called an extreme value distribution if there exists a distribution $F(\mathbf{x})$ and normalizing functions $\mathbf{a}(n) = (a_1(n), \cdots, a_d(n)) \geq 0$ and $\mathbf{b}(n) = (b_1(n), \cdots, b_d(n)) \in \mathbb{R}^d$ such that

$$F^n(\mathbf{a}(n)\mathbf{x} + \mathbf{b}(n)) \to G(\mathbf{x}) \tag{4.1}$$

for $\mathbf{x} \in \mathbb{R}^d$ where $G(\mathbf{x}) \neq 0$, which is equivalent to supposing there exists i.i.d. random vectors $\mathbf{X}_n = (X_{n1}, \cdots, X_{nd}), n \geq 1$ such that

$$P[\bigvee_{i=1}^{n} X_{i1} \leqslant a_1(n)x_1 + b_1(n), \cdots, \bigvee_{i=1}^{n} X_{id} \leqslant a_d(n)x_d + b_d(n)] \to G(\mathbf{x}).$$

When relation (4.1) holds, we say $F \in MDA(G)$. We always assume the one dimensional marginals of G are non-degenerate and we choose the normalising constants such that the j-th ($j = 1, \cdots, d$) marginal G_j of G is a one dimensional extreme value distribution with representation $exp(-(1+\xi x)^{-1/\xi})$ for some $\xi \in \mathbb{R}$.

There is a positive measure ν, associated with G and defined on subsets of $\mathbb{D} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$ such that

$$-log\ G((\mathbf{x}^\xi - 1)/\xi) = -\nu([\mathbf{0}, \mathbf{x}]^c). \tag{4.2}$$

The measure ν is called the exponent measure (see e.g., Resnick [34]) and is finite on compact sets, that is subsets bounded away from $\mathbf{0}$. We have equation (4.1) equivalent to vague convergence of measures on $[\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$

$$nP[(\mathbf{X}_1 - \mathbf{b}(n))/\mathbf{a}(n) \in ((\cdot)^\xi - 1)/\xi] \to \nu(\cdot),$$

where ν satisfies relation (4.2), and this in turn is equivalent to convergence of a sequence of point processes to a limiting Poisson Process (Resnick [34], page 154).

The distribution G satisfies the following property

$$G^t(((t\mathbf{x})^\xi - 1)/\xi) = G((\mathbf{x}^\xi - 1)/\xi) \ , \quad t > 0, \mathbf{x} > \mathbf{0}.$$

This translates to the measure ν which in turn has the following important scaling property (c.f. e.g., de Haan *et al.* [9]) :

$$\nu(sA) \ = \ s^{-1}\nu(A) \ , \quad s > 0, \tag{4.3}$$

where *A* are Borel subsets of $\mathbb{D}$.

We make a polar coordinate transformation. We pick a norm $|\cdot|$ on $\mathbb{R}^2$. Then from equation (4.3) we get

$$\nu\{\mathbf{X} \in \mathbb{D} \ : \ |\mathbf{X}| > x, \mathbf{X}/|\mathbf{X}| \in \Lambda\} \ = \ x^{-1} \, \nu\{\mathbf{X} \in \mathbb{D} : |\mathbf{X}| > 1, \mathbf{X}/|\mathbf{X}| \in \Lambda\} =: \ x^{-1} \, P_\Theta(\Lambda),$$

where Λ is a subset of the unit sphere and where P_Θ is known as the *spectral measure*. In other words, after transformation to polar coordinates, the measure ν is a product measure. The relationship between the spectral measure, P_Θ, and ν is given by

$$\nu(|\mathbf{X}| > 1, \mathbf{X}/|\mathbf{X}| \in \Lambda) \ = \ P_\Theta(\Lambda). \tag{4.4}$$

It is customary, but not obligatory, to use the Euclidean L_2 norm, to parameterize the unit sphere by angles in $[0, \pi/2]$ and to think of the P_Θ as a distribution on subsets of $[0, \pi/2]$. Intuitively, the density of P_Θ evaluated at $\theta \in [0, \pi/2]$ can then be interpreted as a measure of the likelihood that with $\mathbf{X} = (X, Y)$, $arctan(Y/X) \ = \ \theta$ provided that at least one of the components of $\mathbf{X}$ was extreme. Most of the mass concentrating around $\pi/4$ indicates a high correlation between extreme movements of the two components while most of the mass around 0 and $\pi/2$ corresponds to independent occurences of extreme movements in the two components.

In order to estimate the extremal dependence, we therefore estimate the spectral measure which is equivalent to estimating ν.

4.1.1 Estimation of the Spectral Measure

A natural estimator of ν, the tail empirical measure

$$\widehat{\nu}_n(\,.\,) := k^{-1}\sum_{i=1}^{n}\epsilon_{(\mathbf{X}_i-\widehat{\mathbf{b}}(\frac{n}{k}))/\widehat{\mathbf{a}}(\frac{n}{k})}\left(((\cdot)^{\widehat{\xi}}-1)/\widehat{\xi}\right) \tag{4.5}$$

is provided by De Haan and Resnick ([7]).

The following result holds (see Proposition 2.3 in their paper): *Suppose (4.1) holds and* $k(n)\to\infty$, $k(n)/n\to 0$, $k(n)/logn\to\infty$. *Suppose* $\widehat{\mathbf{a}}(\frac{n}{k})$, $\widehat{\mathbf{b}}(\frac{n}{k})$, $\widehat{\xi}$ *are estimators of* $\mathbf{a}(\frac{n}{k})$, $\mathbf{b}(\frac{n}{k})$, ξ *satisfying*

$$\frac{\widehat{\mathbf{a}}(\frac{n}{k})}{\mathbf{a}(\frac{n}{k})}\to 1,\quad \frac{\widehat{\mathbf{b}}(\frac{n}{k})-\mathbf{b}(\frac{n}{k})}{\mathbf{a}(\frac{n}{k})}\to 0,\quad \widehat{\xi}\to\xi$$

a.s. Then $\widehat{\nu}_n$, *defined in eq. (4.5), is a strongly consistent estimator of* ν:

$$\widehat{\nu}_n\to\nu,\quad a.s. \tag{4.6}$$

In particular, one can use estimators $\widehat{\mathbf{a}}(\frac{n}{k})$, $\widehat{\mathbf{b}}(\frac{n}{k})$, $\widehat{\xi}$ given in Dekkers, Einmahl and de Haan [10] and defined below.

Denote the sample by $\{\mathbf{X_i}=(X_{i1},\cdots,X_{id}), i=1,\cdots,n\}$ and indicate the order statistics of the jth components $\{X_{ij},\cdots,X_{nj})$ by

$$X_{(1),j}\leqslant X_{(2),j}\leqslant\cdots\leqslant X_{(n),j}$$

Define the functions

$$\bar{t}=0\wedge t,$$

$$\rho_1(t)=\frac{1}{1-\bar{t}}\ ,\quad \rho_2(t)=\frac{2}{(1-\bar{t})(1-2\bar{t})}\ ,$$

and define the estimators by $(j = 1, \cdots, d)$

$$H_j = k^{-1} \sum_{i=0}^{k-1} logX_{(n-i),j} - logX_{(n-k),j} \ ,$$

$$M_j = k^{-1} \sum_{i=0}^{k-1} (logX_{(n-i),j} - logX_{(n-k),j})^2,$$

$$\widehat{\xi}_j = H_j + 1 - \frac{1}{2}(1 - \frac{H_j^2}{M_j})^{-1},$$

$$\widehat{b}_j(\frac{n}{k}) = X_{(n-k),j} \ ,$$

$$\widehat{a}_j(\frac{n}{k}) = \frac{X_{(n-k),j}\sqrt{3H_j^2 - M_j}}{\sqrt{3(\rho_1(\widehat{\xi_j}))^2 - \rho_2(\widehat{\rho_j})}}.$$

The spectral measure can now be estimated with the help of the tail empirical measure as follows,

$$\widehat{P}_\theta(\ .\ ) := \widehat{\nu}_n(|\mathbf{X}| > 1, \mathbf{X}/|\mathbf{X}| \in (\cdot)) \quad .$$

We interpret $\widehat{P}_\theta$ as a distribution on subsets of $[0, \pi/2]$ and its density evaluated at $\theta \in [0, \pi/2]$ can then be interpreted as a measure of the likelihood that $arctan(\widetilde{X}_2/\widetilde{X}_1) = \theta$ provided that at least one of the components of $(\widetilde{X}_1, \widetilde{X}_2)$ was extreme where

$$(\widetilde{X}_1, \widetilde{X}_2) = ((1+\widehat{\xi}_1((X_1 - \widehat{b}_1(\frac{n}{k}))/\widehat{a}_1(\frac{n}{k})))^{1/\widehat{\xi}_1}, (1+\widehat{\xi}_1((X_2 - \widehat{b}_2(\frac{n}{k}))/\widehat{a}_2(\frac{n}{k})))^{1/\widehat{\xi}_2}) \ . \tag{4.7}$$

Choosing the right value for *k*

A good choice of the parameter k in the tail empirical measure (eq. 4.5) is a prerequisite. Stărică [37] uses the scaling property of ν in eq. (4.3) to help decide on the optimal k. Since $\widehat{\nu}_n$ is a strongly consistent estimator of ν in the limit, the scaling property also holds for $\widehat{\nu}_n$ asymptotically. Defining the set A as

$A = \{(s,t) : s \geq 0, t \geq 0, s^2 + t^2 > 1\}$ and with u varying in the neigbourhood of 1, a good k is therefore such that the following relation holds

$$\widehat{\nu}_n(uA) \approx u^{-1}\widehat{\nu}_n(A) \; . \tag{4.8}$$

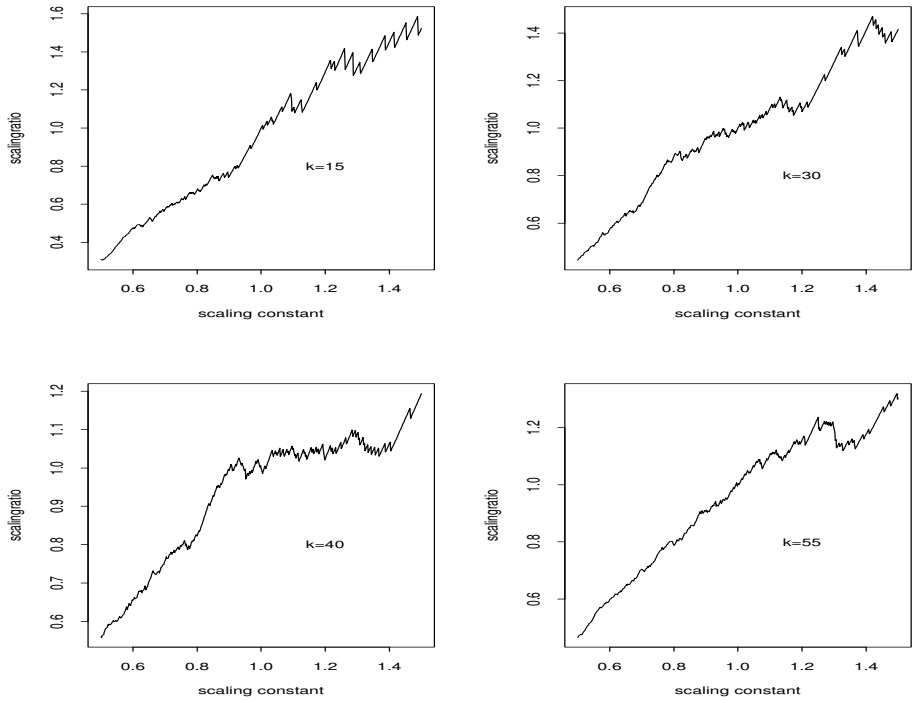

Figure 4.1: *The scaling law for different values of k.*

This method of choosing k is devised by Stărică [37] and is illustrated in Figure 4.1, where the ratio $\widehat{\nu}_n(uA)/u^{-1}\widehat{\nu}_n(A)$ is plotted against u for different values of k. As u decreases, the set uA gets larger, resulting in the counting of points towards the centre of the distribution by ν_n. This in turn causes the failure of the scaling law (4.3) and the violation of (4.8). But for values of u in the neighbourhood of 1, an appropriate choice of k is one which gives a graph which hovers around 1. The upper two graphs illustrates the case where the k values are too small, resulting in us going too far in the tails and hence wasting useful information. The fourth graph shows the case where the k value is too big.

Chapter 5

Regression Model I

In this chapter, we fit Regression Model I (refer to Section 3.2) to two sets of data, namely:

1. Data set 1: Time series of daily returns obtained from contemporaneous foreign exchange noon buying rates in New York City for the British Pound (GBP) and the German Mark (DEM). The data sets range from January 11, 1979 to January 24, 2002.

2. Data set 2: Time series of daily returns calculated using the closing prices of the Standard & Poors 500 Composite Stock Price Index (S&P500) and the Dow Jones Industrial Average Index (DJI). The data sets range from January 03, 1977 to March 06, 2003.

We refer to Data set 1 as GBP/DEM and to Data set 2 as S&P500/DJI. Since the results for the two series are quite similar, for ease of illustration, we mainly show graphs obtained using GBP/DEM but the tabulated results from the statistical tests are provided for both series of data.

5.1 Volatility Estimation

The preliminary step in fitting Regression Model I is to subtract the estimated mean from the log-returns. We then proceed to apply the two-sided exponential kernel regression (eq. 3.2) to the square zero-mean returns with bandwidth

$h = 120$ for GBP/DEM and bandwidth $h = 145$ for S&P500/DJI to estimate the volatilities.

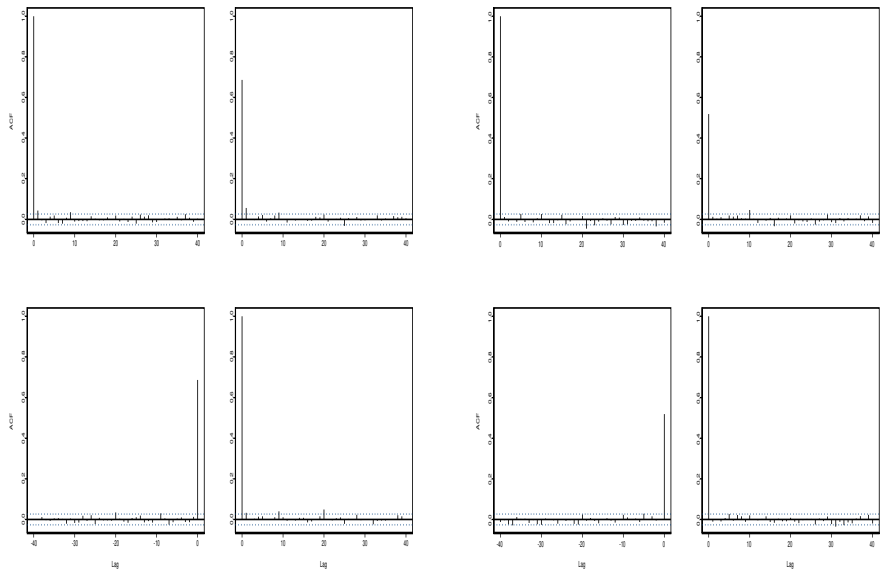

Figure 5.1: Left four plots: *SACFs/SCCFs for the estimated residuals ($\widehat{\varepsilon}_t$) obtained by standardization with the changing variances.* Right four plots: *SACFs/SCCFs for ($|\widehat{\varepsilon}_t|$).* Note: *For each, the plot at the intersection of row i and column j corresponds to the SACF/SCCF of the coordinate i and past lags of the coordinate j, where $i, j = 1, 2$ and $1, 2$ corresponds to GBP and DEM respectively.*

We choose the bandwidth parameters such that the long range dependence aspect of the SACFs and SCCFs of the absolute returns disappears. We can judge if a good choice has been made by looking at the autocorrelation structure of the estimated residuals ($\widehat{\varepsilon}_t$) and their absolute values ($|\widehat{\varepsilon}_t|$) as well as the Ljung-Box test plots. Under the null hypothesis

$$H_o : \widehat{\varepsilon}_t \quad \text{are independent vectors,}$$

we expect the SACFs and the SCCFs to stay within the boundary of the 95%-confidence intervals for most lags, as they do in Figure 5.1, hence supporting the

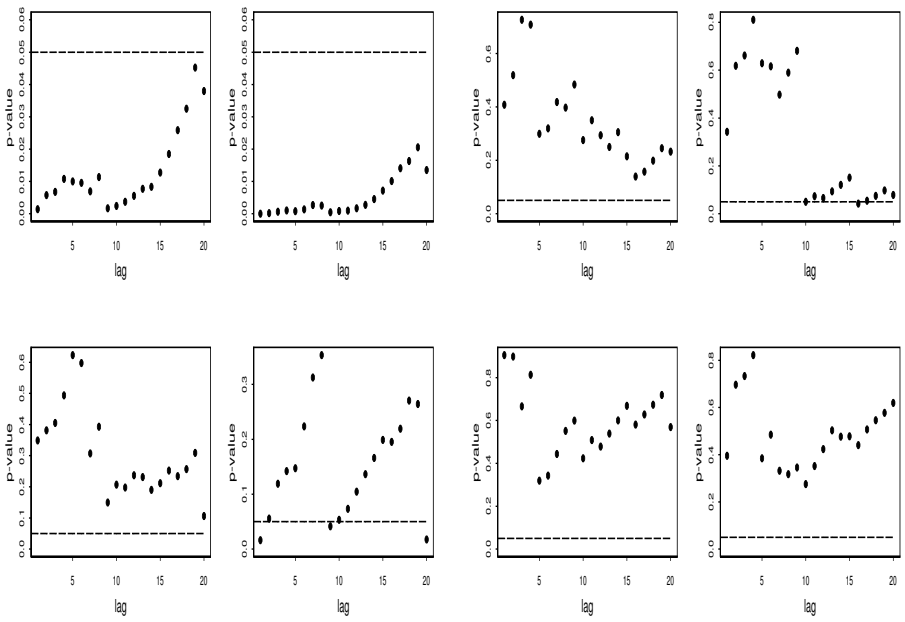

Figure 5.2: Left four plots: *Ljung Box test p-values at different lags for the estimated residuals ($\widehat{\varepsilon}_t$) obtained by standardization with the changing variances.* Right four plots: *Ljung Box test p-values at different lags for ($|\widehat{\varepsilon}_t|$).* Note: *For each, the plot at the intersection of row i and column j is based on the SACF/SCCF of the coordinate i and past lags of the coordinate j, where $i, j = 1, 2$ and $1, 2$ corresponds to GBP and DEM respectively. The broken line has been drawn at a p-value of .05.*

null hypothesis at a significance level of $.05$.

There are formal techniques for choosing the bandwidth such as cross-validation but we did not look at them here.

However, the plot of the p-values from the Ljung-Box test in Figure 5.2 indicates the presence of serial correlation in the GBP residuals and the presence of serial cross-correlation between the GBP and DEM estimated residuals for most lags. On the other hand, the Ljung-Box test based on the SACFs and SCCFs of the absolute residuals accept the hypothesis of independence for most of the first 20 lags as seen in right plots of Figure 5.2.

Periods	GBP/DEM			S&P500/DJI		
	1	2	3	1	2	3
Lower value	0.65	0.79	0.54	0.62	0.54	0.37
Upper value	0.70	0.82	0.61	0.67	0.60	0.44

Table 5.1: *Checking the hypothesis that the contemporaneous dependence between the coordinates of the standardized residuals ($\widehat{\varepsilon}_t$) do not change across the three periods. The lower and upper values of the* 95% *confidence interval for the standard Pearson sample correlation coefficient are given.*

Drees and Stărică [12] come across a similar situation and offer a possible explanation for it. They argue that dividing by the changing volatility changes only the absolute values and not the signs of the returns. As a result, we see a change in the SACF of the absolute residuals but not much in that of the residuals. The latter behaviour may indicate either weak dependence in the signs of the returns or non-stationarities in the mean. In the model though, only the non-stationarities that might be present in the second moment structure of the time series are taken care of.

Figures 5.3, 5.4, 5.5 and 5.6 show the returns, the estimated unconditional standard deviation and the standardized residuals for the GBP/DEM and S&P500/DJI data sets . The heteroskedasticity seems to be quite reduced in the residuals compared to the original series of returns.

5.2 Subperiod analysis

In order to check if the residuals are identically distributed, we split the residuals in three subsamples of equal length.

First, we estimate a 95% confidence interval for the standard Pearson sample correlation coefficient ρ between the coordinates for each subsample. The results are tabulated in Table 5.1.

The marginals are found to be dependent for both series of data. The clear lack of overlapping between the confidence intervals for the three periods indicates that the dependence between the coordinates changes with time for both data sets.

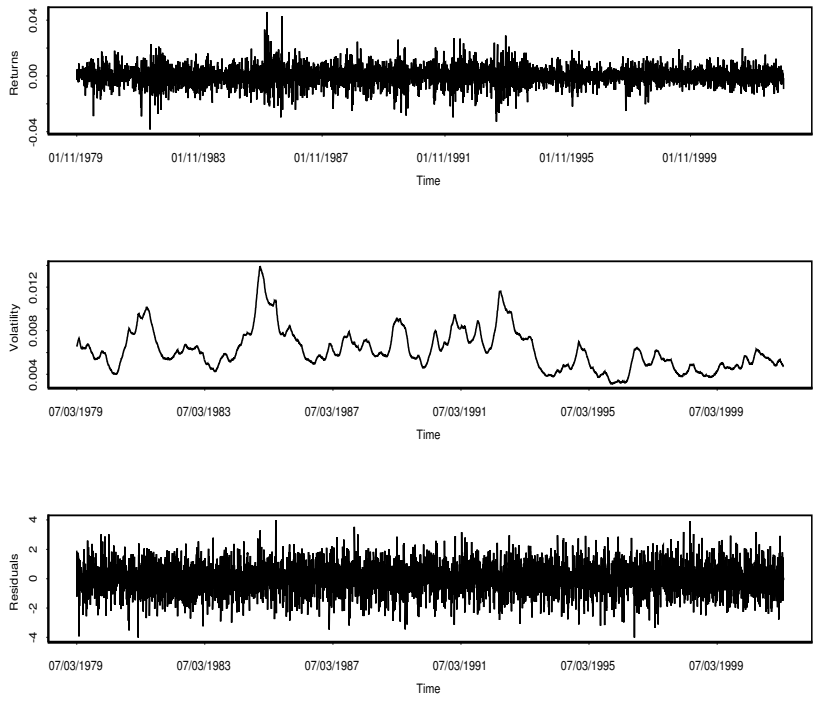

Figure 5.3: Top plot: *GBP returns*. Middle plot: *The corresponding volatility process* $\widehat{\sigma}(t)$ *estimated using the exponential kernel estimator*. Bottom plot: *The standardized residuals.*

Secondly, for each coordinate, we perform a pairwise comparison of the three empirical distribution functions using the Kolmogorov-Smirnov test[1]. Table 5.2 shows that at the .01 significance level, we reject the null hypothesis that the marginal distribution of the first coordinate does not change between period 1 and period 2 for both series of data. For the second coordinate, the null hypothesis is rejected across all three periods at the .05 significance level by the S&P500/DJI data. For all other periods and for both coordinates, the null hypothesis is not rejected at the .05 significance level.

Thus, even though the residuals showed very little evidence of serial depen-

[1]Kolmogorov-Smirnov is not strictly applicable since there may remain some dependence in the data. However, we judge this to be of minor importance.

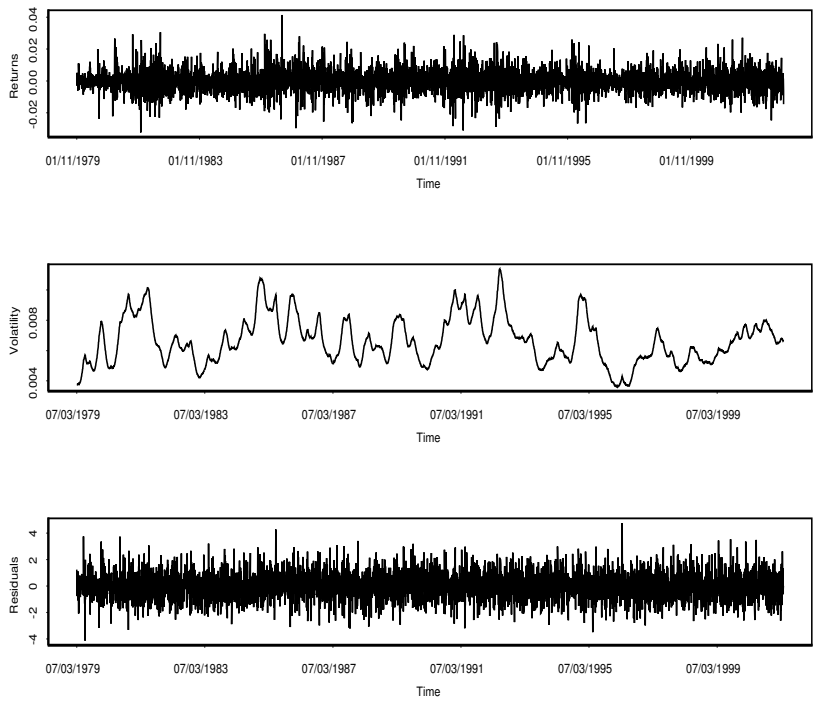

Figure 5.4: Top plot: *DEM returns*. Middle plot: *The estimated volatility process* $\widehat{\sigma}(t)$. Bottom plot: *The standardized residuals.*

dence, they do not seem to be identically distributed.

In our next step, we investigate the extremal dependence between the coordinates.

5.3 Extremal Dependence

We concentrate on the extremal dependence in the left tail. For that purpose, we divide the residuals into three subsamples of equal length. As from here, we then work with the contemporaneous positive values of the negated residuals or, equivalently, with the contemporaneous coordinates from the lower left quadrant.

We check the tails of their marginal distributions by calculating the confidence intervals around the Maximum Likelihood Estimate (MLE) for the shape parame-

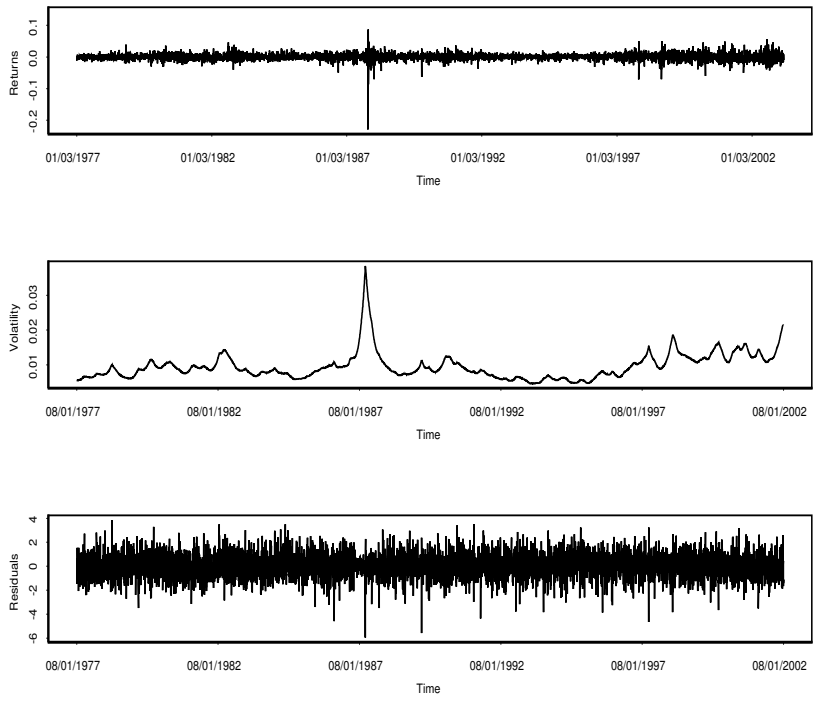

Figure 5.5: Top plot: *S&P*500 *returns*. Middle plot: *The estimated volatility process* $\widehat{\sigma}(t)$. Bottom plot: *The standardized residuals.*

ter ξ of the Generalized Pareto Distribution (GPD) (see Appendix B) fitted to the scaled excesses over a high threshold.

The overlap in the 95% confidence intervals for ξ (refer to Tables 5.3 and 5.4) show that the marginal tail shape parameters for the GBP/DEM and S&P500/DJI data are about the same for the three periods and for each of the coordinates. The ξ values are indicative of light tails.

We measure the extremal dependence with the help of the *spectral measure* (see relation 4.4). The latter is estimated using the tail empirical measure (refer to eq. 4.5). The values of k to be used in its estimation are determined according to the method used by Stărică [37], as explained in Section 4.1.1.

The scaling ratio at these values for each of the three periods are plotted in Figure 5.7 (see Left plots) for the GBP/DEM data.

Marginals	GBP/DEM			S&P500/DJI		
	Periods			Periods		
	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
1	.004	.14	.18	.001	.077	.20
2	.37	.57	.098	.026	.005	.044

Table 5.2: *Checking the hypothesis that the marginal distributions of the standardized residuals ($\widehat{\varepsilon}_t$) do not change across the three periods . Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test and the p-values are given.*

Marginals	GBP/DEM		
	Periods		
	1	2	3
1	$(-0.314, 0.109)$	$(-0.456, 0.160)$	$(-0.390, -0.004)$
2	$(-0.249, 0.167)$	$(-0.482, 0.056)$	$(-0.369, 0.075)$

Table 5.3: 95% *confidence interval for the MLE of the shape parameter ξ of the GPD for the marginals of the contemporaneous standardized residuals ($\widehat{\varepsilon}_t$) from the lower left quadrant for the GBP/DEM data.*

Marginals	S&P500/DJI		
	Periods		
	1	2	3
1	$(-0.372, -0.034)$	$(-0.156, 0.409)$	$(-0.269, 0.182)$
2	$(-0.479, -0.043)$	$(-0.260, 0.236)$	$(-0.169, 0.485)$

Table 5.4: 95% *confidence interval for the MLE of the shape parameter ξ of the GPD for the marginals of the contemporaneous standardized residuals ($\widehat{\varepsilon}_t$) from the lower left quadrant for the S&P500/DJI data.*

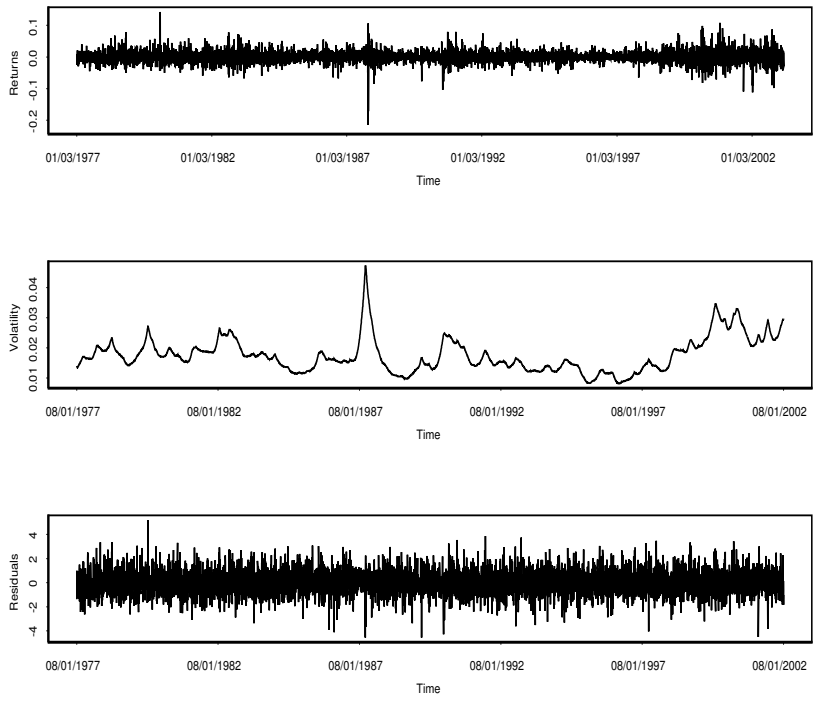

Figure 5.6: Top plot: *DJI returns*. Middle plot: *The estimated volatility process* $\widehat{\sigma}(t)$. Bottom plot: *The standardized residuals.*

In order to investigate if the extremal dependence changes across the three periods, we work with the standardized residuals scaled and transformed according to equation (4.7) for each coordinate at each period. The contemporaneous pairs of scaled residuals whose distance from the origin is greater than a threshold value of 1 are considered as *extreme* (see eq. 4.4) . The angles at which these *extreme* pairs are located with respect to the x-axis are then calculated and denoted by $\theta_1, \theta_2, \theta_3$ for the three periods respectively.

Estimated densities of the spectral measures (see eq. 4.4) for the GBP/DEM residuals are illustrated through the histogram/density plots of θ_1, θ_2 and θ_3 (see Right plots of Figure 5.7).

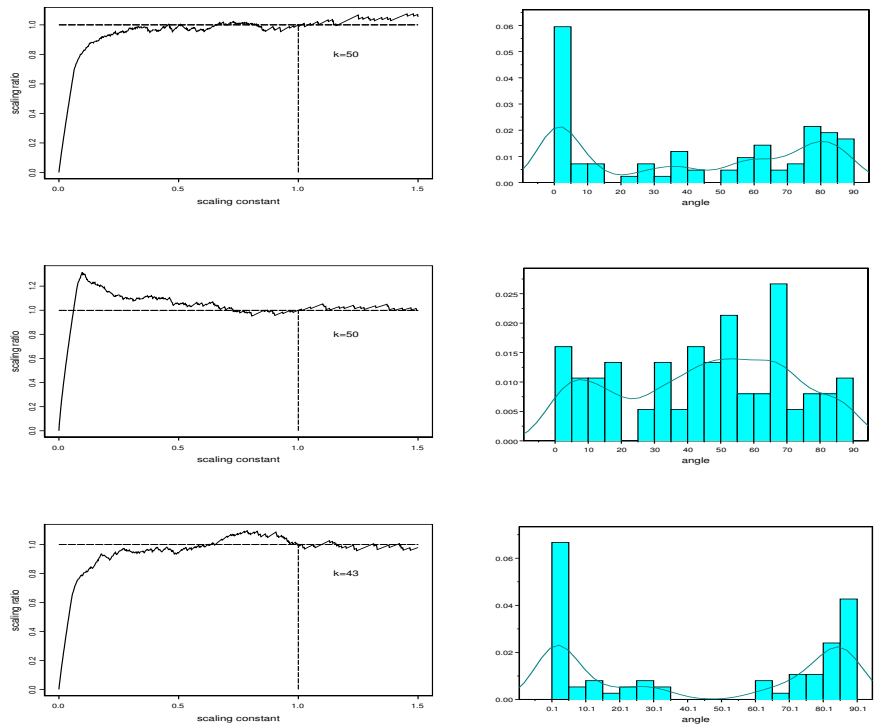

Figure 5.7: *The scaling law for* $k = 50$ *(top and middle) and* $k = 43$ *(bottom) and the estimated density/histogram of the spectral measure for the GBP/DEM contemporaneous standardized residuals* ($\widehat{\varepsilon}_t$) *from the lower left quadrant for period* 1 *(top),* 2 *(middle) and* 3 *(bottom) respectively.*

5.3.1 Non-Parametric Testing

In this section, we check the hypothesis that $\theta_1, \theta_2, \theta_3$ belong to the same population by doing a pairwise comparison of the empirical distribution functions of the vectors of angles with the help of the Kolmogorov-Smirnov test and the New Rank Test, a non-parametric test outlined in the Appendix A.

The results from Table 5.5 show that for the GBP/DEM data, the hypothesis of identical distributions is rejected by both tests for period 2 versus 3 and by the New Rank Test for period 1 versus 2 at a significance level of .01. The results for the S&P500/DJI data show that the hypothesis is rejected at the .05 significance level by both tests for periods 1 versus 2 and by the Kolmogorov-Smirnov test for

Periods	GBP/DEM			S&P500/DJI		
	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
K-S	.028	.21	.003	.039	.24	.028
N-R	$p < .01$	$.05 < p$	$p < .01$	$.01 < p < .05$	$.05 < p$	$.05 < p$

Table 5.5: *Checking that the angle populations are identical. Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test (K-S) and the New Rank test (N-R).*

period 2 versus 3.

5.3.2 Conclusion

The results from this chapter allow us to conclude that the residuals standardized by the changing variances can be modeled as independent vectors but that the contemporaneous dependence changes through time. The distribution of the marginals is characterized by changes in the center and lack of change in the tail estimates. As for the extremal dependence between the residuals, it is found to change across time.

The implications of such findings are that accounting for the changing variance is enough to remove the linear temporal dependence observed in the absolute returns but there is a need for the changing covariance structure to be taken into account.

Chapter 6

Regression Model II

In this chapter we fit Regression Model II (see Section 3.3) to the GBP/DEM and S&P500/DJI data sets. For ease of illustration, we again provide the graphs only for the GBP/DEM data set.

6.1 Covariance Estimation

We first substract the estimated mean from the log-returns and then apply a two-sided kernel regression (see eq. 3.4) to the resulting zero-mean return series to estimate the covariance matrix. Since it is hard to find the common parameter which works as well in estimating the variance as the covariance simultaneously, we have used the same parameter as for the variance.

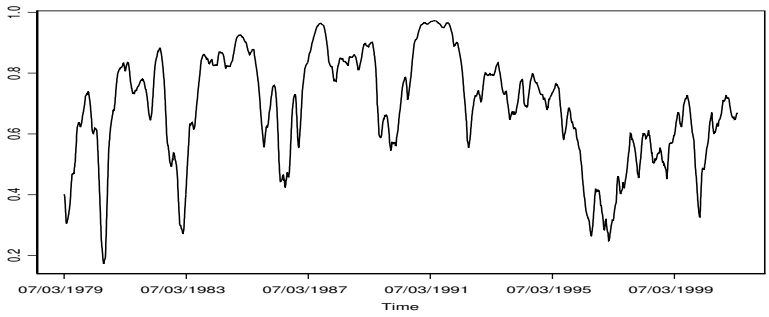

Figure 6.1: *Plot of the correlation between the GBP and DEM returns estimated using the exponential kernel regression estimator.*

The residuals are given by

$$\widehat{\epsilon}_t = \widehat{\mathbf{S}}^{-1}(t)(\mathbf{r}_t - \widehat{\mu}_t), \quad t = b+1, \cdots, n-b,$$

where $\widehat{\mathbf{S}}(t)$, the square root of the estimate of the covariance matrix $\mathbf{S}^2(t)$, is calculated using the eigenvalue decomposition method.

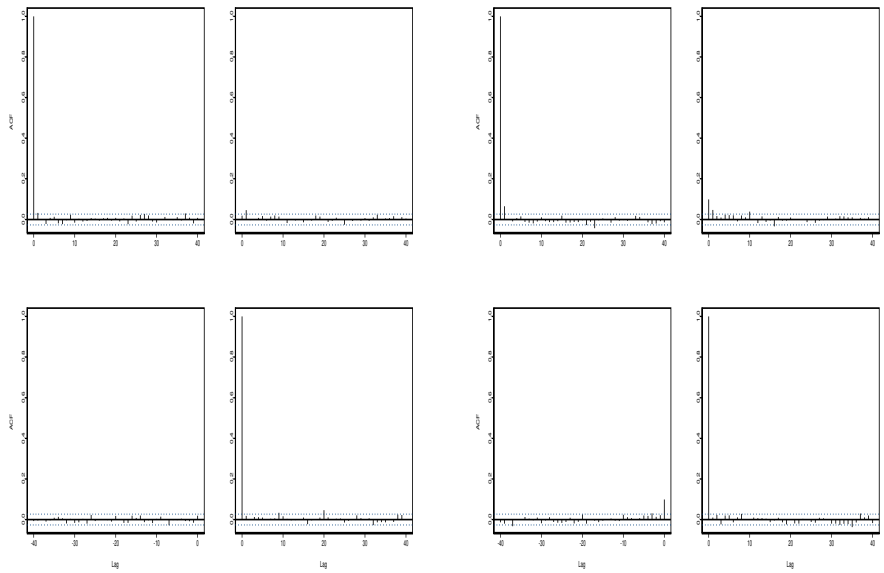

Figure 6.2: Left four plots: *SACFs/SCCFs for the estimated GBP/DEM residuals ($\widehat{\epsilon}_t$) obtained by standardization with the changing covariance.* Right four plots: *SACFs/SCCFs for ($|\widehat{\epsilon}_t|$).*

The estimated correlation between the two sets of data is plotted in Figure 6.1.

A look at the autocorrelation structure of the estimated residuals ($\widehat{\epsilon}_t$) and their absolute values ($|\widehat{\epsilon}_t|$) (see Figure 6.2) indicates that the SACFs and SCCFs stay within the boundary of the 95%-confidence intervals, drawn under the null hypothesis of i.i.d. residuals, for most lags, hence supporting the null hypothesis that $\widehat{\epsilon}_t$ are independent vectors at a significance level of $.05$.

This is further supported by the plot of the p-values from the Ljung-Box test

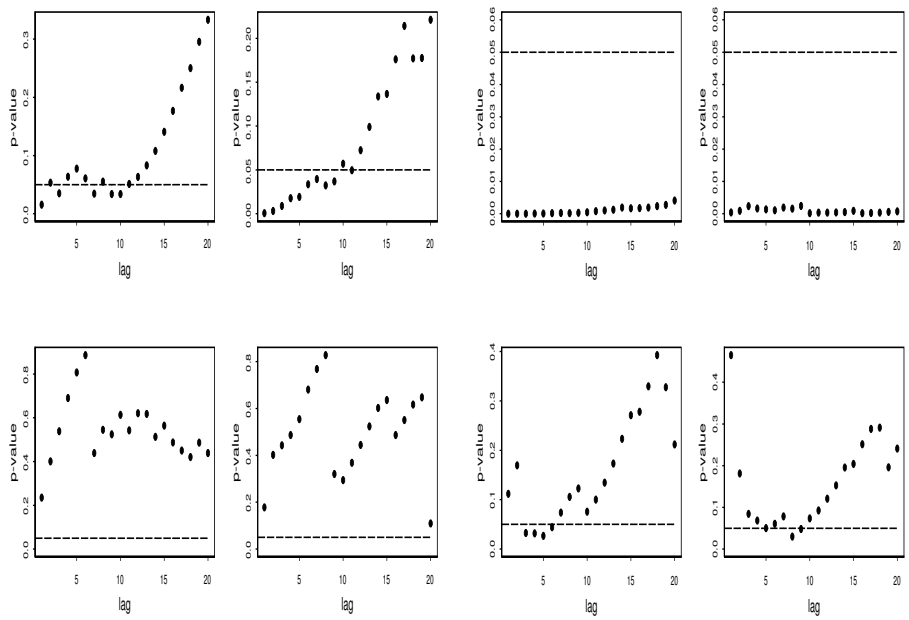

Figure 6.3: Left four plots: *Ljung Box test p-values for different lags for the estimated GBP/DEM residuals ($\widehat{\epsilon}_t$) obtained by standardization with the changing covariance.* Right four plots: *Ljung Box test p-values for different lags for ($|\widehat{\epsilon}_t|$).* Note: *For each, the plot at the intersection of row i and column j is based on the SACF/SCCF of the coordinate i and past lags of the coordinate j, where $i, j = 1, 2$. The broken horizontal line has been drawn at a p-value of .05.*

(see Appendix A) for the residuals in Figure 6.3. On the other hand, the presence of serial correlation in the first coordinate of the absolute residuals and the presence of serial cross-correlation between the coordinates of the absolute residuals can be observed from the Ljung-Box test plots in Figure 6.3.

One possible explanation for this situation could be due to the fact that the chosen parameters estimated the variance properly but did not perform as well for the covariance. Or it could be that the model did not fit perfectly.

Periods	GBP/DEM			S&P500/DJI		
	1	2	3	1	2	3
Lower value	-0.03	-0.01	-0.04	-0.04	-0.05	-0.05
Upper value	0.06	0.08	0.05	0.05	0.04	0.04

Table 6.1: *Checking that the contemporaneous dependence between the coordinates of the covariance standardized residuals ($\widehat{\epsilon}_t$) does not change across the three periods. The lower and upper values of the 95% confidence interval for the standard Pearson sample correlation coefficient for the three periods.*

Marginals	GBP/DEM			S&P500/DJI		
	Periods			Periods		
	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
1	.038	.33	.070	.044	.28	.62
2	.99	0.19	0.44	.24	.84	.24

Table 6.2: *Checking that the marginal distributions of the covariance standardized residuals ($\widehat{\epsilon}_t$) do not change across the three periods . Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test and the p-values are given.*

6.2 Subperiod Analysis

The results from Tables 6.1 and 6.2 further seem to support the null hypothesis of i.i.d. residuals with independent marginals for both the GBP/DEM and the S&P500/DJI data series.

The overlap in the 95% confidence intervals indicates that the dependence between the coordinates stays constant across time. The fact that the confidence intervals contain the zero is also an indication that the coordinates are uncorrelated. As for the marginal distributions of the standardized residuals, they are found to change across period 1 and 2 for one of the marginals for both data series at a significance level of .05 but not at a significance level of .01.

So far, we have seen that accounting for the changing covariance seems to take care

Marginals	GBP/DEM		
	Periods		
	1	2	3
1	$(-0.228, 0.213)$	$(-0.356, 0.149)$	$(-0.196, 0.389)$
2	$(-0.219, 0.191)$	$(-0.293, 0.179)$	$(-0.375, 0.086)$

Table 6.3: 95% *confidence interval (C.I.) for the MLE of the shape parameter* ξ *of the GPD for the marginals of the contemporaneous covariance standardized residuals* $(\widehat{\epsilon}_t)$ *from the lower left quadrant for the GBP/DEM data.*

Marginals	S&P500/DJI		
	Periods		
	1	2	3
1	$(-0.355, 0.112)$	$(-0.118, 0.380)$	$(-0.315, 0.130)$
2	$(-0.403, -0.006)$	$(-0.318, 0.136)$	$(-0.149, 0.373)$

Table 6.4: 95% *confidence interval for the MLE of the shape parameter* ξ *of the GPD for the marginals of the contemporaneous covariance standardized residuals* $(\widehat{\epsilon}_t)$ *from the lower left quadrant for the S&P*500*/DJI data.*

of the linear dependence. The null hypothesis of identically distributed residuals has also been supported by the statistical tests carried out. In the next section, we use extreme value theory to check for possible non-linear dependence. More specifically, we investigate the stationarity of the dependence structure between the coordinates across time, concentrating mainly on the *extremes*.

6.3 Extremal Dependence

We again concentrate on the extremal dependence in the left tail and follow the same procedure as in the previous chapter.

Estimates of the left tails of the residuals (see Tables 6.3 and 6.4) indicate that they are quite stable across time and for each of the marginals for both the GBP/DEM and the S&P500/DJI data sets.

The values of k to be used in the estimate of the spectral measure for the GBP/DEM residuals at the three different periods are determined to be $52, 45$ and

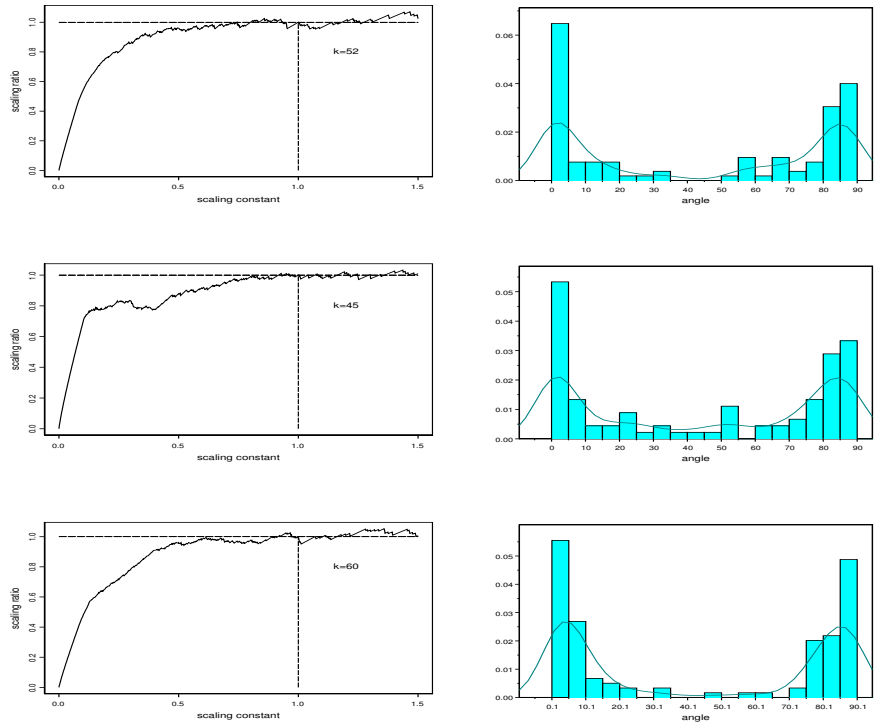

Figure 6.4: *The scaling law for* $k = 52$ *(top),* $k = 45$ *(middle) and* $k = 60$ *(bottom) and the estimated density/histogram of the spectral measure for the GBP/DEM contemporaneous covariance standardized residuals from the lower left quadrant for period* 1 *(top),* 2 *(middle) and* 3 *(bottom) respectively.*

60 respectively. The scaling ratio at these values as well as the corresponding estimated densities of the spectral measures are plotted in Figure 6.4 (Left plots). See Figure 6.5 for the density plots corresponding to the S&P500/DJI data.

In all of the three periods, we see that most of the mass is concentrated around the axes indicating extremal independence.

6.3.1 Non-Parametric Testing

In this section, we perform the pairwise comparison of the empirical distribution functions of the angles corresponding to the positions of the extreme residual pairs. We check that $\theta_1, \theta_2, \theta_3$ belong to the same population. The results from Ta-

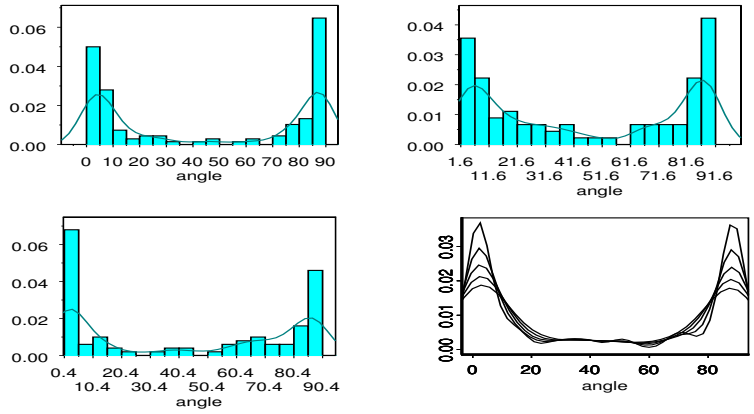

Figure 6.5: Top Plots and Bottom Left Plot: *The estimated density/histogram of the spectral measure for the S&P*500*/DJI contemporaneous covariance standardized residuals from the lower left quadrant for period* 1*,* 2 *and* 3. Bottom Right: *Estimated density curves of the spectral measure for the S&P*500*/DJI contemporaneous covariance standardized residuals from the lower left quadrant for the whole period. The curves correspond to a family of smoothed histograms with different binwidths.*

ble 6.5 indicate that the extremal dependence stays constant at the .05 significance level across the three periods for the GBP/DEM data set. The null hypothesis of identical distribution is rejected at the .01 significance level by the New Rank Test for the S&P500/DJI data between period 2 and 3 but accepted at the same level for all the other periods and by both tests.

6.4 Whole period analysis

Since accounting for the changing covariance results in i.i.d. residuals with stationary extremal dependence, we can carry out the same analysis on the whole sample of bivariate returns.

Estimated densities of the spectral measure for the whole sample for the GBP/DEM (see Figure 6.6) and for the S&P500/DJI (see bottom right plot of Figure 6.5) residuals show most of the mass being concentrated around the axes. This result indi-

Periods	GBP/DEM			S&P500/DJI		
	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
K-S	.67	.089	.11	.18	.28	.021
N-R	$.05 < p$	$.05 < p$	$.05 < p$	$.05 < p$	$.01 < p < .05$	$p < .01$

Table 6.5: *Checking that the angle populations are identical. Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test(K-S) and the New Rank test (N-R).*

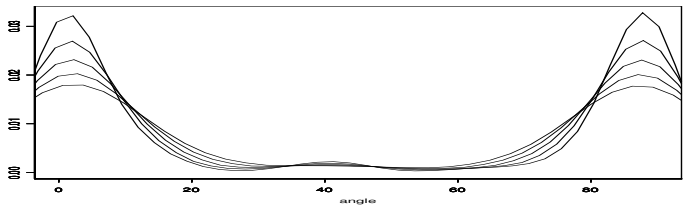

Figure 6.6: *Estimated density curves of the spectral measure for the GBP/DEM contemporaneous covariance standardized residuals from the lower left quadrant for the whole period. The curves correspond to a family of smoothed histograms with different binwidths.*

cates extremal independence between the coordinates which corroborate with the expectations we had following the subperiod analysis.

6.5 Conclusion

In this chapter, we fitted Regression Model II to our data sets and estimated the covariance matrix of the bivariate returns data using the two-sided exponential kernel regression estimator.

The results from the analysis of the dependence between the coordinates as well as the marginal distributions of the resulting residuals across time were much improved compared to the results from Regression Model I so that the null hypothesis of identically distributed residuals was better supported. The extremal dependence between the coordinates of the transformed residuals was also found to stay reasonably constant across time. Since the resulting residuals also seemed

independent in the extremes, it indicated that most of the extremal dependence could be due to the changing covariance structure of the returns. And accounting for this changing covariance is enough to go from non-stationary bivariate returns to residuals whose extremal dependence is stationary.

Chapter 7

GARCH Modelling

As an alternative to the non-parametric approach used in the last two chapters in estimating the volatility, we now turn to a conditional parametric modelling of the volatility, namely the GARCH models. We restrict attention to the GARCH (1,1) model.

All the computations in this chapter have been carried out on the GBP/DEM and the S&P500/DJI data sets using the GARCH module in Splus. We show the graphs only for GBP/DEM since the results are very similar for the S&P500/DJI data.

7.1 Univariate GARCH

We start by fitting a univariate GARCH (1,1) model with constant mean to the univariate returns separately. We thus have the following model for the returns

$$r_t = \mu + X_t,$$

where r_t is the returns, μ is the mean and

$$X_t = \sigma_t Z_t,$$

where

$$\sigma_t^2 = \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2.$$

We used the heavy-tailed t-distribution for the conditional distribution of X_t given the past in the likelihood since, as we have observed earlier when looking

at the stylized features of financial data, returns tend to have heavier tails than the Gaussian distribution. By default, the number of degrees of freedom for the t-distribution is estimated as part of the maximum likelihood fitting procedure in the GARCH module.

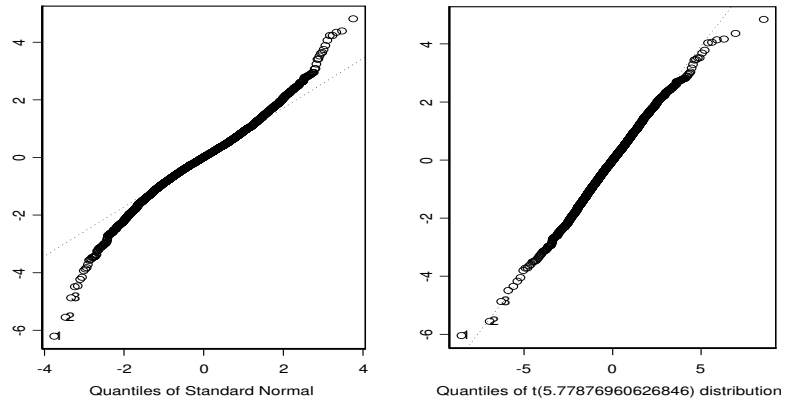

Figure 7.1: *QQ-plots of the GBP GARCH (1,1) standardized residuals*. Left: *Conditional Gaussian distribution used*. Right: *Conditional t-distribution used*.

As an illustration, see Figure 7.1, where the QQ-plots of the standardized residuals are drawn when the Gaussian distribution is used as conditional distribution (see Left plot) and when the t-distribution with estimated 5.78 degrees of freedom is used (see Right plot). A better fit is obtained with the t-distribution while the left plot indicates that the standardized residuals have a somewhat heavy-tailed deviation from a Gaussian distribution in the tails.

The ACF of standardized residuals and of their absolute values show very little evidence of dependence as shown in Figure 7.2.

On the other hand, subperiod analysis of the standardized residuals indicates a change in the marginal empirical distributions across time at the $.01$ significance level, especially for the S&P500/DJI data, while the marginals are more stable for the GBP/DEM data. The dependence between the coordinates is also found to clearly change with time as shown by the complete lack of overlapping between the confidence intervals for the standard Pearson sample correlation coefficient across

Periods	GBP/DEM			S&P500/DJI		
	1	2	3	1	2	3
Lower value	0.63	0.79	0.54	0.61	0.56	0.40
Upper value	0.68	0.82	0.60	0.66	0.61	0.47

Table 7.1: *Checking that the contemporaneous dependence between the coordinates of the univariate GARCH (1,1) standardized residuals do not change across the three periods. The lower and upper values of the* 95% *confidence interval for the standard Pearson sample correlation coefficient for the three periods.*

Marginals	GBP/DEM			S&P500/DJI		
	Periods			Periods		
	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
1	.12	.12	.053	.002	.12	.066
2	.23	.60	.041	.005	.004	.015

Table 7.2: *Checking that the marginal distributions of the univariate GARCH (1,1) standardized residuals do not change across the three periods . Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test and the p-values are given.*

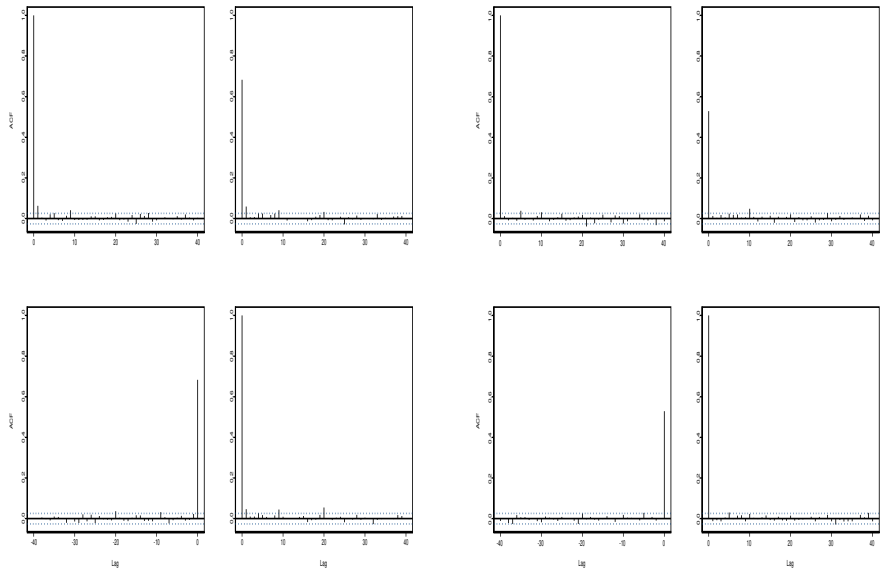

Figure 7.2: *SACFs/SCCFs for the univariate GARCH (1,1) standardized GBP/DEM residuals (Left four plots) and for their absolute values (Right four plots).*

the three periods. The results are tabulated in Tables 7.1 and 7.2 and allow us to conclude that an assumption of i.i.d. residuals is not reasonable.

We therefore need to take into account more than just the changing conditional variance and do so by fitting a bivariate GARCH (1,1) model to the bivariate returns.

7.2 Bivariate GARCH Model

The model we use in this case is

$$\mathbf{r}_t = \mu + \mathbf{X}_t,$$

where all the quantities in the above equation are vectors of dimension 2. $\mathbf{r}_t$ is the bivariate vector of returns, μ is the vector of mean values of $\mathbf{r}_t$. The conditional

distribution of $\mathbf{X}_t$ given the past is a bivariate *t*-distribution with the conditional covariance matrix, denoted by V_t. We choose to work with the BEKK (1,1) model, introduced by Engle and Kroner [14] and which has the following conditional covariance matrix structure

$$V_t = AA^T + A_1(X_{t-1}X_{t-1}^T)A_1^T + B_1V_{t-1}B_1^T.$$

The presence of a paired transposed matrix factor for each of the 2x2 parameter matrices A, A_1 and B_1 ensures symmetry and non-negative-definiteness of the conditional covariance matrix though it should be noted that A, A_1, B_1 need not be symmetric.

The ACF of the standardized residuals and of their absolute values show very little evidence of dependence as shown in Figure 7.3.

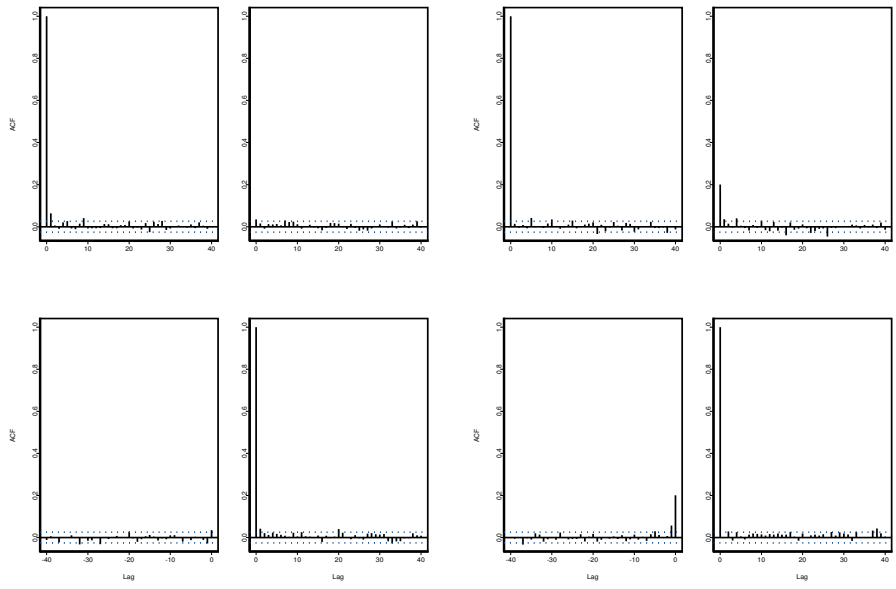

Figure 7.3: *SACFs/SCCFs for the bivariate GARCH (1,1) standardized GBP/DEM residuals (Left four plots) and for their absolute values (Right four plots).*

	GBP/DEM			S&P500/DJI		
	Periods			Periods		
Marginals	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
1	.11	.041	.034	.003	.080	.037
2	.15	.080	.0009	.63	.12	.070

Table 7.3: *Checking that the marginal distributions of the bivariate GARCH (1,1) standardized residuals do not change across the three periods . Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test and the p-values are given.*

	GBP/DEM			S&P500/DJI		
Periods	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
K-S	0.023	0.002	0.025	0.21	0.070	0.061
N-R	$.01 < p < .05$	$p < .01$	$p < .01$	$.05 < p$	$.05 < p$	$p < .01$

Table 7.4: *Checking that the angle populations using the Bivariate GARCH (1,1) model are identical. Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test (K-S) and the New Rank test (N-R).*

In this case, the confidence intervals for the standard Pearson sample correlation coefficient overlapped across the three periods implying that the contemporaneous dependence between the marginals was more stable across time.

The marginal distributions were more stable across time (see Table 7.3) for the S&P500/DJI residuals compared to the univariate GARCH modelling case. On the other hand, we see less stability for the marginals of the GBP/DEM residuals. In the univariate GARCH case, they changed only across period 2 and 3 for the second marginal at the .05 level. They also change here though we reject the null hypothesis this time at the .01 level. But the null hypothesis of identical marginals is also rejected at the .05 level across periods 1 and 3 and periods 2 and 3 by the first marginal.

We move next to the extremal dependence between the coordinates of the residuals in the left tail. The results from the New Rank Test in Table 7.4 indicate that the extremal dependence changes at the .01 significance level between period 1 and

3 for the GBP/DEM data as well as between period 2 and 3 for both the GBP/DEM and the S&P500/DJI data.

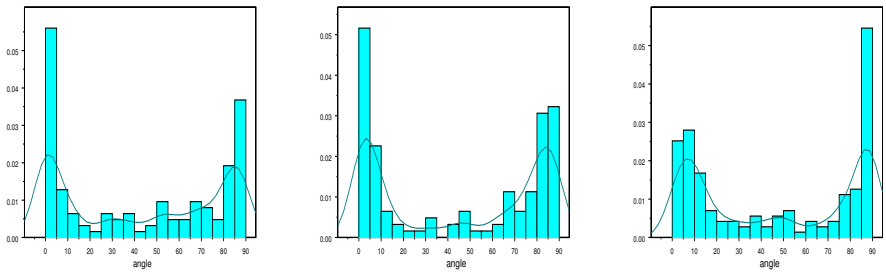

Figure 7.4: *The estimated density/histogram of the spectral measure for the bivariate GARCH (1,1) contemporaneous GBP/DEM residuals from the lower left quadrant for period* 1 *(Left),* 2 *(Middle) and* 3 *(Right).*

The estimated densities of the spectral measures are plotted in Figure 7.4 for the GBP/DEM data.

They do not look so different across the three periods even though the statistical tests indicate otherwise. In all of the three periods, we also see that most of the mass is concentrated around the axes indicating extremal independence.

To see if fitting different models would give different results, we fitted first a different bivariate GARCH (1,1) model and then more complicated BEKK models, namely the BEKK (1,2), BEKK (2,1) and BEKK (2,2) to the GBP/DEM data. We also experimented with other conditional distributions than the t-distribution.

From the point of view of the Box-Test Plots for the absolute residuals, the BEKK (1,1) model with the conditional t-distribution seemed to perform best in filtering out the heteroskedasticity.

7.3 Conclusion

In this chapter, we first filtered out the changing conditional volatilities from the bivariate returns using univariate GARCH (1,1) models. The standardardized residuals were not found to be i.i.d.

The bivariate GARCH (1,1) model seemed to do a good job in filtering out

the heteroskedasticity. The S&P500/DJI residuals were found to be closer to i.i.d. and with more or less stable extremal dependence across time. On the other hand, less stability was observed for the marginals of the GBP/DEM residuals and the statistical tests indicated less stable extremal dependence across time. However, visual inspection of the estimated spectral densities indicated a lot of similarity across the three periods and also pointed to extremal independence between the marginals.

Chapter 8

Direct Extreme Value Analysis

In this chapter, we perform a direct extreme value analysis of the GBP/DEM and the S&P500/DJI returns data sets using the spectral measure theory[1]. The extreme bivariate coordinates are then declustered so as to be made more independent.

In the univariate context, essentially two approaches have been proposed for declustering (refer to Embrechts et al. [15] for more details). The first one identifies events as clusters of high-level exceedances and filters out the cluster peaks. Specifically a *model threshold* and a *separation interval* are chosen such that if two exceedances over the model threshold are closer together than the separation interval then they are deemed part of the same cluster. If the time interval between successive exceedances is longer than the separation interval, it indicates the end of the old cluster and the beginning of a new one. The second approach is instead based on dividing the time into blocks and extracting the maximum value within each block. An improvement of the first methodology is proposed by Smith [36] and Davison and Smith [6] .

Shortly, there is no universally accepted method for identifying clusters and even simple-minded approaches seem to work just as well in practice. Usually rough intuitive judgement is used in the choice of the separation interval and it is recommended that different values be used for comparison.

The situation is a bit more complicated in the multivariate context. Coles and Tawn [5] achieve multivariate declustering by fixing a large enough time block

[1]The tail empirical measure is still a consistent estimator of the spectral measure under the type of dependency structure commonly used in the econometric time series modelling (see Stărică [37]).

length and independent events are filtered out by concatenating the maxima of the marginal processes within each block. Nadarajah [29] applies univariate declustering to the marginal processes at low threshold levels and concatenate independent extremes in each margin with the concurrent values in the other margins to produce independent events.

In the sequel, we follow an approach closer to the second one since it results in contemporaneous coordinates and also does not produce events which are not actually realized. In that respect, it allows us to compare our results with the ones obtained using the previous models in the previous chapters. However, our analysis differs in various aspects.

We first filter out the extreme values and then identify independent events from them. In so doing, we end up with contemporaneous coordinates and the extreme bivariate coordinates extracted, using the norm, are those which are extreme in either of the margins.

Secondly, Nadarajah [29] filters out the marginals into cluster peaks according to the separation interval and then concatenate them with concurrent values in the other margins. We do not work with marginal cluster peaks but with all the large values in each margin.

Thirdly, Nadarajah declusters even further within the concatenated coordinates by grouping those events which are still closer than the separation interval and choosing a single event out of these according to three possible rules:

1. by choosing the one that is closest to the central time of the grouped event;
2. at random among the events;
3. by choosing the one corresponding to the most extreme value of the structure variable of interest.

We instead keep track of the time indexes which the extreme coordinates correspond to in the original data. A separation interval is then chosen. If the time interval between successive pairs of coordinates is longer than the separation interval, they are treated as being effectively independent. Otherwise the coordinates are filtered for rejection.

	GBP/DEM			S&P500/DJI		
Periods	1 vs 2	1 vs 3	2 vs 3	1 vs 2	1 vs 3	2 vs 3
K-S	0.24	.48	0.058	0.041	.53	0.11
N-R	$.05 < p$	$.05 < p$	$.01 < p < .05$	$.01 < p < .05$	$.05 < p$	$.05 < p$

Table 8.1: *Checking that the angle populations for the declustered data are identical. Pairwise comparison of the periods are performed using the Kolmogorov-Smirnov test (K-S) and the New Rank test (N-R).*

A separation interval of 5 days was chosen. Due to the already scarce number of extreme values, we minimised the rejection level by rejecting only if the *accumulated* separation time was less than 6. As an illustration, suppose that three bivariate coordinates have time indexes 18, 20 and 24. We reject the second value only and not the third one, even though the difference between them is only 4 days. The remaining two bivariate coordinates will then have time indexes 18 and 24 and can be considered to be effectively independent. Under the block method or the method adopted by Nadarajah [29], these three would belong to the same cluster and we would end up with only one bivariate coordinate instead of two. So in effect, we optimise the use of the already scarce data at this extreme level.

8.1 Data Analysis Results

For purposes of comparison, we divided the data into three periods and followed the same procedure as we did in Section 5.3. Here, we worked with the returns directly instead of the residuals.

The extreme pairs of coordinates were then chosen using the declustering method outlined above so as to make them more independent. The angles at which the resulting *independent* extreme coordinates were located with respect to the x-axis were then calculated and the null hypothesis that they belonged to the same population tested.

The end results are tabulated in Table 8.1 and show that the extremal dependence remains stable across time at the .01 significance level for both data sets. Visual inspection of the estimated densities of the spectral measure (see Bottom

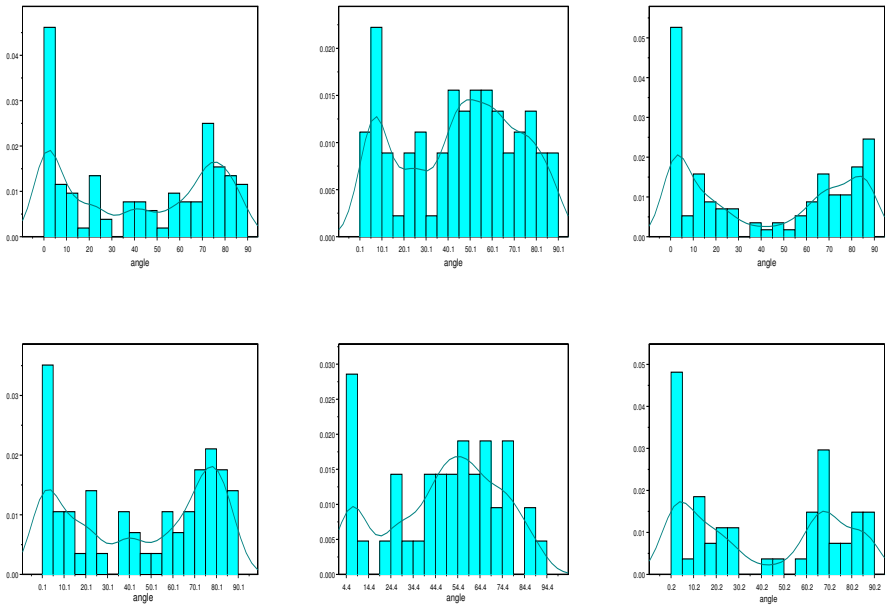

Figure 8.1: *The estimated density/histogram of the spectral measure for the non-declustered (Top Row) and declustered (Bottom Row) GBP/DEM data for period* 1 *(Left),* 2 *(Middle) and* 3 *(Right).*

Rows of Figures 8.1 and 8.2) however indicate relatively more dependence in the extremes in the second period for the declustered S&P500/DJI data and even more for the declustered GBP/DEM data.

A comparison of the estimated densities of the spectral measure for the non-declustered data (see Top Row of Figures 8.1 and 8.2) shows that the shape of the density curves is almost the same as for the declustered data.

8.2 Conclusion

In this chapter, we carried out a direct investigation of the extremal dependence between the returns data. The extreme bivariate coordinates were chosen so as to make them more independent before comparing their extremal dependence across time.

The end results indicated that the extremal dependence between the declus-

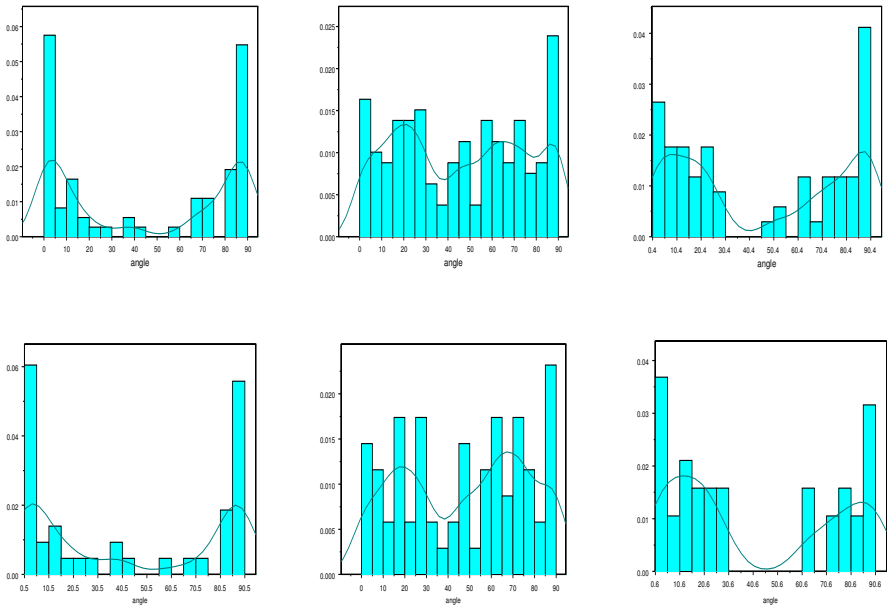

Figure 8.2: *The estimated density/histogram of the spectral measure for the non-declustered (Top Row) and declustered (Bottom Row) S&P*500*/DJI data for period* 1 *(Left),* 2 *(Middle) and* 3 *(Right).*

tered extreme coordinates remained constant across time at a significance level of .01. The histograms/density plots of the angles however indicated relatively more extremal dependence in the second period than in the first and third periods.

Chapter 9

Conclusion

In this thesis, we discussed different methodologies of analyzing bivariate returns with a specific focus on the behaviour of the extremes.

We used a non-stationary, unconditional non-parametric approach to model in a first attempt the changing variances and then in a second step the changing covariance structure. The second methodology used the stationary, conditional parametric approach of GARCH modelling in order to fit the changing conditional volatility. Thirdly, we performed a direct investigation of the extremal dependence between the returns data. We proposed a declustering method to make the extreme bivariate coordinates more independent before comparing their extremal dependence across time. The goodness of fit on data of the first two methods used was checked through a detailed investigation of the evolution of the marginal distribution of residuals as well as the central and extremal dependence between the coordinates.

To exemplify the use of the three different approaches, we analyzed two sets of data. The first one consisted of daily returns obtained from contemporaneous foreign exchange noon buying rates in New York City for the British Pound (GBP) and the German Mark (DEM). The second pair of data consisted of daily returns calculated using the closing prices of the Standard & Poors 500 Composite Stock Price Index (S&P500) and the Dow Jones Industrial Average Index (DJI).

Under the non-parametric non-stationary framework, the residuals standardized by the changing unconditional variances were practically uncorrelated vectors with dependent coordinates. The linear dependence was found to change with time.

An analysis of the extremal dependence between the coordinates of the residuals showed that it also varied through time, pointing towards the need for a changing covariance structure to be taken into account. Similar results were obtained when filtering out the changing conditional variances using the univariate GARCH model.

When a changing covariance structure was estimated under the non-parametric non-stationary framework, the residuals were close to being independent and identically distributed. The extremal dependence between the coordinates of the residuals remained quite stable across time and was close to independence.

In the parametric and stationary framework, filtering out the conditional covariance using the bivariate GARCH model gave similar results for the S&P500/DJI data. On the other hand, mixed evidence was found in the case of the GBP/DEM data set. Statistical tests indicated changes in the extremal dependence while visual inspection of the estimated densities of the spectral measure seemed to confirm their stability across time.

In the case of the direct extreme value analysis of the returns, we found that the extremal dependence between the declustered extreme coordinates seemed to remain constant across time.

The methodologies we applied seemed to perform well in allowing us to draw conclusions about the extremal dependence for bivariate returns. The limited evidence provided in this work suggests that the dynamics of bivariate returns is characterized by changes of the covariance structure. The fact that the residuals appear to be independent in the extremes could imply that a large part (if not all) of the extremal dependence in bivariate returns could be explained by the changing linear relationship captured by the covariance. However, more indepth statistical analysis is needed in order to confirm this statement. This will be the subject of future research.

Bibliography

[1] Baumgartner, W., Weiß, P., and Schindler, H. (1998). A Nonparametric Test for the General Two-Sample Problem. *Biometrics*, **54**, 1129-1135.

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

[3] Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *J. Econometrics*, **52**, 5-59.

[4] Bollerslev, T., Engle, R., F. and Nelson, D., B. (1994). ARCH Models. In R.F. Engle and D.L. McFadden (Eds.). *Handbook of Econometrics*, Volume IV, Chapter 49. Elsevier Science B.V.

[5] Coles, S., G. and Tawn, J., A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc.: Series B*, **53**, 377-392.

[6] Davison, A., C. and Smith, R., L. (1990). Model for exceedances over high thresholds. *J. Roy. Statist. Soc.: Series B*, **52**, 393-442.

[7] de Haan, L. and Resnick, S. (1993). Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models*, **9(2)**, 275-309.

[8] de Haan, L.,Jansen,D.,Koedijk, K. and De Vries, C. (1994). *Safety first portfolio selection, extreme value theory and long run asset risks*, Galambos, J.(ed.), *Extreme Value Theory and Applications*, Kluwer, 471-487.

[9] de Haan, L. and de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes*, **1**, 7-46.

[10] Dekkers, A., Einmahl, J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833-1899.

[11] Diebold, F. X., Schuermann,T. and Stroughair, J. D. (2000). Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance*, **1**, 30-36.

[12] Drees, H. and Stărică, C. (2002). A simple non-stationary model for stock returns. *Available at http://www.math.chalmers.se/ starica.*

[13] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimate of the Variance of UK Inflation. *Econometrica*, **50**, 987-1008.

[14] Engle, R. F. and Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, **11**, 122-150.

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

[16] http://www.federalreserve.gov/releases/

[17] Ghysels, E., Harvey, A. C. and Renault E. (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, 119-191. Amsterdam: North-Holland.

[18] Guillaume,D., M., Dacorogna, M., M., Davé, R., D., Müller, U. A., Olsen, R., B., and Pictet,O. V. (1997). From the Bird's Eye to the Microscope - a survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, **1**, 95-129.

[19] Hsu, D., Miller, R. and Wichern, D. (1974). On the stable paretian behaviour of stock-market prices. *J. Amer. Statist. Assoc.*, **69**, 108-113.

[20] Jansen, D., W. and De Vries, C., G. (1991). On the frequency of Large Stock Returns: Putting Booms and Bursts into Perspectives. *Review of Economics and Statistics*, **73**, 18-24.

[21] Ljung, G., M. and Box, G., E., P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297-303.

[22] Longin, F. and Solnik, B. (1995). Is the Correlation in International Equity Returns Constant: 1960-90? *J. Int. Money Finance*, **14**, 3-26.

[23] Longin, F. and Solnik, B. (2000). Extreme Correlation of International Equity Markets: 1960-90? *Journal of Finance*, **56**, 2, 649-676.

[24] Loretan, M. and Phillips, P.C.B. (1994). Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets. *Journal of Empirical Finance*, **1**, 211-248.

[25] Mikosch, T. (2002). Modelling dependence and tails of financial time series via regular variation. *Semstat Meeting in Gothenburg in December 2001*.

[26] Mikosch, T. and Stărică, C. (2002). Changes of structure in financial time series and the GARCH model. Available at http://www.math.chalmers.se/ starica.

[27] Mikosch, T. and Stărică, C. (2002). Long range dependence effects and ARCH modelling. In: Boukhan, P., Oppenheim, G. and Taqqu, M.S. (EDs.). Long Range Dependence. Birkhauser, Boston. To appear.

[28] Mikosch, T. and Stărică, C. (2002). Non-stationarities in financial in financial time series, the long range dependence and the IGARCH effects. Available at http://www.math.chalmers.se/ starica.

[29] Nadarajah, S. (2001) Multivariate declustering techniques. *Environmetrics*, **12**, 357-365.

[30] Nadaraya, E. (1964). On Estimating Regression. *Theory Prob. Appl.*, **10**, 186-190.

[31] Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, **59**, 347-370.

[32] Pagan, A., R. and Schwert, G., W. (1990). Testing for covariance stationarity in stock market data. *Economics Letters*, **33**, 165-170.

[33] Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal Appl. Probab.*, **8**, 745-756.

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

[35] Shephard, N. (1996). Statistical Aspects of ARCH and Stochastic Volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models - in econometrics, financa and other fields*, pp. 1-67. London: Chapman Hall.

[36] Smith, R., L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications*, pp. 621-638, editor J. Tiago de Oliveira, Reidel, Dordrecht.

[37] Stărică, C. (1999). Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, **6**, 515-553.

[38] Sprent, P. (1993). *Applied Nonparametric Statistical Methods*, 2nd edition. Chapman & Hall, London.

[39] Taylor, S., J. (1986). *Modelling Financial Time Series*. Wiley, Chichester.

[40] Taylor, S., J. (1994). Modelling Stochastic Volatility: A Review and Comparative Study. *Mathematical Finance*, **4**, 183-204.

[41] Watson, G. (1964). Smooth Regression Analysis. *Sankhya, Series A*, **26**, 359-372.

Appendix A

Statistical Tests for the Models

Ljung-Box test (Ljung and Box [21])

Under the null hypothesis of independence in a time series, the Ljung-Box test statistic $T(n)$ follows asymptotically a χ^2-distribution with *n* degrees of freedom, with $T(n)$ being calculated as

$$T(n) = L(L+2)\sum_{i=1}^{n}\frac{{\rho_i}^2}{L-i},$$

where ρ_i is the sample auto- or cross- correlation function at lag *i* and L is the length of the time series.

Non-parametric testing:

In order to test the null hypothesis that two sets of data come from the same distribution, we compare the empirical distribution functions using the non parametric tests described below.

For the first data set of mutually independent variables in increasing order of magnitude, $X_1, X_2, \cdots, X_n$, the empirical distribution is defined by

$$F_n(x) = \begin{cases} 0, & \text{for } x < X_1 \\ i/n, & \text{for } X_i \le x < X_{i+1} \\ 1, & \text{for } x > X_n \end{cases}$$

$F_m(x)$ is defined accordingly for the sample $Y_1, Y_2, \cdots, Y_m$.

1. **Kolmogorov-Smirnov Test**

   The test statistic is the maximum of $|F_n(x) - F_m(x)|$ multiplied by a proportionality factor. It is a powerful test if the variances of the two populations are different (Baumgartner [1]).

2. **New Rank Test** (Baumgartner [1])

   The new rank test is the integral of the square norm of the difference $(F_n(x) - F_m(x))$, weighted by its variance. This emphasizes the tails of the distribution functions. It is at least as powerful as other standard nonparametric tests like the Kolmogorov-Smirnov and the Cramer-Von Mises Tests (Baumgartner [1]) but is very useful if the alternative hypothesis is not known, since it is independent of whether the populations differ in their means, variances, or only in the shapes. For sample sizes as small as 5, the asymptotic distribution approximates the exact one quite well and that holds even if $n \neq m$. The critical values are 2.493 at a significance level of 5% and 3.880 at a significance level of 1%.

Appendix B

The Peaks Over Thresholds(POT) Method

As mentioned in the beginning of Chapter 4, the numerous approaches by which extreme values may be statistically modelled separate into two forms: methods for maxima over fixed intervals and methods for exceedances over high thresholds. We introduce very briefly some theory about the second method here.

The appropriate limit theory in this context is based on a point process result of Pickands [33]. The limit result suggests modelling exceedances of a high threshold by a non-homogeneous Poisson process. A consequence of this model is that the excess values over the threshold follow the Generalized Pareto distribution (GPD) and that maximum values are modelled by the Generalized Extreme Value Distribution, both distributions having a common shape parameter ξ.

The key idea of this approach is explained below.

Suppose

$$X, X_1, \cdots, X_n \quad \text{are i.i.d. with } dfF \in MDA(G),$$

where G is the Generalized Extreme Value Distribution with parameters (μ, σ, ξ). Then, for large enough u, the distribution function of $(X-u)|X > u$ can be approximated by a member of the family

$$H(y) = 1-(1+\xi\frac{y}{\sigma(u)})^{-1/\xi}, \quad \text{defined on} \quad \{y : y > 0 \text{ and } 1+(\xi y)/\sigma(u) > 0\} \tag{B-1}$$

The family of distributions defined by eq. B-1 is called the generalized Pareto family. It has two parameters. ξ is the shape parameter and coincides with the shape parameter in the GEV model and $\sigma(u)$ is the scale parameter.

A key modelling aspect here is the selection of the threshold. Too low a threshold is likely to violate the asymptotic basis of the model, leading to bias, while too high a threshold will result in few excesses with which the model can be estimated, leading to high variance. One method used in selecting the threshold is in comparing the stability of parameter estimates based on the fitting of the models at different thresholds. Then above a level at which the asymptotic motivation for the GPD is valid, estimates of the shape parameter ξ should be approximately constant. Sampling variability means that the estimates of ξ will not be exactly constant, but they should be stable with respect to their sampling errors.

For more details on the POT method, refer to Embrechts et al. [15].