Doctoral Thesis ETH No. 15283

Copula Inference for Finance and Insurance

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
DOKTOR DER MATHEMATIK

presented by
ALEXANDRA DA COSTA DIAS
M.Sc. Math. ISEG - UTL Portugal
born July 31, 1967
citizen of Portugal

accepted on the recommendation of
Prof. Dr. P. Embrechts, examiner
Prof. Dr. A. J. McNeil, co-examiner
Prof. Dr. M. Ivette Gomes, co-examiner

2004

Institutional support:
This work was developed at the Department of Mathematics of ETH Zurich with financial support from: Grant SFRH/BD/993/2000 from Fundação para a Ciência e a Tecnologia - FCT/POCTI and Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal.

à memória do meu pai

Acknowledgment

A large number of people contributed directly and indirectly to the accomplishment of this thesis. Most of all, I want to thank my supervisor Prof. Paul Embrechts. He introduced me to the subject of my work and gave me the opportunity that conducted to this thesis. His support and contagious enthusiasm was absolutely essential for me. I thank Prof. Andrade e Silva, Prof. Lourdes Centeno and Prof. Egídio dos Reis from ISEG, Lisbon for introducing me to Prof. Paul Embrechts. I also would like to thank Prof. Alexander J. McNeil and Prof. M. Ivette Gomes for accepting to be co-examiners of this thesis.

I am grateful to Prof. Alexander J. McNeil, Filip Lindskog, Prof. Guus Balkema and Dr. Wolfgang Breymann for several fruitful discussions. I own a special thank to Mrs. Baltes who helped me at ETH and kept remembering me that there are other interesting things to do in Switzerland besides working. I take pleasure in thanking those at the Department of Mathematics of the Faculty of Sciences and Technology of the New University of Lisbon who supported this project during my stay at ETH Zurich.

At ETH I made good friends who helped me while working on this thesis, especially Roger, Tino, Filip and Hansjörg. From Portugal I enjoyed the support of my family and old friends Carmo and Rui. A very special thank goes to Rui.

Finally, I would like to thank Olsen Data for providing the FX data.

Contents

Abstract

This thesis describes models and inference methods for insurance and finance data by means of copulae. Although the definition of a copula dates from 1959 its merits have long been disregarded in insurance, finance and risk management applications. Only in 1997, Wang proposed copula models for the analysis of insurance data. Embrechts, McNeil and Straumann in 1999 applied the copula framework to financial data. Meanwhile, copula theory has been developed up to the textbook level. Whilst the literature on copulae is devoted either to probabilistic theory, to inference methods or to applications, we present a combination of these three modelling aspects. Our approach has been spurred by the recent developments in integrated risk management of financial institutions, resulting in an increasing need for the modelling of multivariate financial risk factors and their interaction.

In this thesis we study copula models for insurance and finance data together with inference methods that allow for the proper fitting of these models. It is for this reason that the development of the theory and the practical applications go hand in hand. Starting from representative insurance and finance data, we look for appropriate copula models to analyse the data. Where necessary, we first develop the methodological tools to perform the copula analysis. Once these tools are available and sufficiently tested we apply them to answer specific econometric or risk management related questions. We present copula models both for independent and identically distributed data as well as for more general time series. In the case of general time series, we consider time–invariant and time–varying copula models, and further study copula change–point

analysis. Two different inference methods are used, a parametric and a semi–parametric one. We explain the standard error estimation procedure for the model parameters and show how confidence intervals can be derived. Moreover, we investigate asymptotic tail–dependence properties, spectral densities, and extreme tail–dependence copula. Goodness–of–fit and elliptical symmetry tests are also considered.

Kurzfassung

Diese Arbeit vereint Modelle und Inferenzmethoden zur Analyse von Versicherungs- und Finanzdaten mit Hilfe von Copulas. Obschon die Definition des Copula-Begriffs auf das Jahr 1959 zurückreicht, hielt diese Art der statistischen Modellierung im Versicherungs- und Finanzbereich und im Risikomanagement erst vor kurzem Einzug. Die Verwendung von Copula-Modellen für Versicherungsdaten wurde 1997 erstmals von Wang propagiert, während Embrechts, McNeil und Straumann im Jahre 1999 diesen Ansatz auf Finanzdaten übertrugen. Mittlerweile ist die Entwicklung der Copula-Theorie bis zum Lehrbuch-Niveau ausgereift. Während die Literatur zum Thema Copula vorwiegend dreigeteilt ist in entweder wahrscheinlichkeitstheoretische, statistische oder anwendungsorientierte Aspekte, ist es das Ziel dieser Arbeit, eine disziplinenübergreifende Synthese dieser drei Aspekte zu präsentieren. Motiviert wird unser Ansatz durch die jüngsten Entwicklungen im Bereich integriertes Risikomanagement von Finanzdienstleistern, wo die Modellierung von mehrdimensionalen Risikofaktoren und deren wechselwirkung unabdingbar geworden ist.

In dieser Arbeit untersuchen wir Copula-Modelle für Versicherungs- und Finanzdaten. Insbesondere sind wir an der Kalibrierung solche Modelle interessiert. Aus diesem Grunde gehen Präsentation der Modelle und ihre praktischen Anwendungen Hand in Hand. Ausgehend von realen Versicherungs- und Finanzdaten, halten wir Ausschau nach adäquaten Copula-Modellen. Wo nötig, werden entsprechende Modelle zuerst entwickelt. Sind die Modelle erst einmal bereitgestellt und hinreichend getestet, können sie dazu verwendet werden, auf Fragen ökonometrischer Art und aus dem Bereich Risikomanagement Antworten zu geben. Dabei

beschränken wir uns nicht nur auf die Modellierung von unabhängigen und identisch verteilten Daten, sondern richten unser Augenmerk auch auf allgemeinere Zeitreihen. Für den letzteren Fall betrachten wir sowohl zeitunabhängige wie auch zeitabhängige Copula-Modelle sowie die Technik der Wechselpunkte. Wir präsentieren eine parametrische und eine semi-parametrische Inferenzmethode. Ferner erläutern wir, wie die Standard-Fehler der Modellparameter geschätzt werden und zeigen, wie daraus Vertrauensintervalle bestimmt werden können. Darüberhinaus widmen wir uns der asymptotischen Abhängigkeit für das Auftreten gemeinsamer extremer Ereignisse und den Spektral-Dichten. Schliesslich werden auch Anpassungstests und Tests auf elliptische Symmetrie durchgeführt.

Introduction

Within current risk management practice, an efficient modelling of dependent risks has become absolutely necessary in finance and insurance. The theory and the markets urged the economic agents to combine individual components. Financial instruments based on more than one asset, like multi–name options or multi–asset portfolios, are widely used and have to be priced and risk managed. Similarly, insurance products involving several lines of business or combined with financial products, as equity linked life insurances for instance have become standard. Contemporary finance also demands good quantitative management of aggregated risks. As part of a whole, risk factors are seldom independent and can even have rather intricate dependence structures. Therefore, flexible multivariate models and good inference methods are called for.

Formerly, in finance (mainly) and insurance (to some extent) there were two main approaches to the dependence modelling question: independence or multivariate normality. Either the several risks were admitted to be independent like in a portfolio of insurance policies, or multivariate normality was assumed as in the classical analysis of financial time series. In each case, risk aggregation is straightforward but often too far from providing realistic models.

A multivariate model has two components: the univariate (marginal) one, which characterises each of the variables and the dependence structure between these marginal variables. This obvious separation can be specified in mathematical terms and it is exploited for modelling in this thesis. The dependence structure of the random variables is known as the copula.

Through the combination of copulae with specific univariate distributions it is possible to construct an infinite number of multivariate distributions which we will refer to as copula–based models.

The notion of a function characterising the dependence structure between several random variables comes from the work of Hoeffding in the early forties. Independently, other authors introduced related notions afterwards but it was in 1959 that Sklar [78] defined and named as copula a functional that gives the multivariate distribution as a function of the univariate marginal distributions. Allowing for the study of the dependence structure apart from the univariate behaviour of each variable, copulae became very useful. The literature kept growing as the interest in copulae increased, for instance in environmental data modelling. In the late nineties this notion was brought into finance and insurance problems. In 1997, Wang [81] proposes copulae for modelling aggregate loss distributions of correlated insurance policies. Frees and Valdez [35] in 1998 and Klugman and Parsa [52] in 1999 use copulae to model bivariate insurance claim data. Embrechts et al. [29] give a comprehensive overview of the application of copulae to finance in a preprint already available on–line in 1999.

By now, the theory of copulae is well established. On the other hand, the literature on inference questions is still scarce. References like Genest et al. [37], Genest and Rivest [38], Joe [48] and Oakes [66] are important in this context. We can find copula models fitted to insurance losses in Blum et al. [3], Frees and Valdez [35] and Klugman and Parsa [52] and to finance data in Breymann et al. [8], Dias and Embrechts [18, 19], Fortin and Kuzmics [33], Patton [67] and Rockinger and Jondeau [72]. Our interest lies on parsimonious multivariate copula–models, able to capture the relevant facts of finance and insurance data and on the inference methods for those models. Such fitted models are essential to accurately estimate functionals of dependent risks, forecast financial time series and evaluate risk measures.

Outline. This thesis has a dual goal. First, to introduce a series of tools for the statistical analysis of risk management questions in finance and insurance. Second, analyse in full a bivariate tick–by–tick foreign exchange (FX) data set, concentrating in particular on (joint) extremes and sta-

tistical properties across frequencies. A novel consequence of the latter analysis is the discovery of so–called stylised facts of dependence in multivariate financial (FX) data. The methodology introduced is applicable to a variety of applied risk management problems.

This thesis is for a large part based on four research papers. Chapter 1 is devoted to the introduction of most of the basic definitions, properties and methods used in the later chapters. We give the definition of a copula together with important examples for finance and insurance. We introduce relevant properties for copula inference. The inference for margins (IFM) and the pseudo log–likelihood methods for copula fitting are detailed. Properties of the estimators, standard errors and confidence intervals estimation methods are also presented. The main data set analysed is introduced, namely, bivariate tick–by–tick FX spot rates for US dollar quoted against German mark (USD/DEM) and quoted against Japanese yen (USD/JPY).

The stylised facts in financial univariate time series are well–known. They include stochastic volatility, heavy–tailed innovations, extremes and seasonalities. In Chapter 2 we perform a first analysis of the dependence structure of the USD/DEM and USD/JPY log–returns across six different time frequencies from one hour up to one day. The first important step is the data deseasonalisation. Originally the log–returns reveal weekly and intra–day seasonalities that are to be removed. We present a deseasonalisation method and apply it to the high–frequency data set. After that, we fit a set of copula families to the deseasonalised bivariate log–returns and estimate the tail–dependence coefficient. Goodness–of–fit tests are given. We also test for ellipticity, a property which is important in a risk management context. We estimate the spectral densities across frequencies for the tail observations and finally fit copula families to the bivariate excesses of the one hour deseasonalised log–returns. This chapter is based on [8] with some modifications.

The analysis performed in Chapter 2 assumes the bivariate observations to be independent identically distributed (iid). This means that the bivariate stationary dependence structure is being estimated for the different time frequencies. The goal of Chapter 3 is to go one step further to the analysis of the conditional dependence structure of time series observations.

In that way, the observations are not treated any more as iid and time series models have to be fitted to the margins. We use GARCH type processes to model the data heteroscedasticity. After the GARCH modelling, the standardised residuals are extracted and their dependence structure is analysed. To the standardised residuals (or filtered log–returns) we fit copula families for the six frequencies. The tail–dependence coefficients are estimated and ellipticity is tested. For the extremal filtered log–returns we estimate the spectral densities and the extreme tail–dependence copula. The results from this chapter are to be compared with those obtained in Chapter 2. It should be stressed that the copula–based models in this chapter assume a dynamic behaviour of the margins but a time–invariant dependence structure. The natural model refinement is to consider dynamics in the margins as well as dynamic behaviour of the copula. This is pursued in the subsequent chapter. Part of [19] is to be found in this chapter.

In Chapter 4 we check the time–invariant copula assumption of Chapter 3 fitting bivariate matrix–diagonal GARCH models to the deseasonalised log–returns for the six frequencies. These models allow for a time–varying cross–correlation and in fact, the estimates point to a time–varying dependence structure. We propose a bivariate dynamic copula–based model where the margins are of GARCH type and the copula has time varying (dependence) parameters. We assume that the copula family is the same over time but allow the dependence parameters to change from realisation to realisation of the process. Throughout the previous copula analyses, the t–copula comes out as a good model, so in this chapter we end up estimating the dynamics of the correlation (one of the t–copula parameters) over time for the six frequencies. The analysis reveals interesting macro–economic aspects underlying the data. The time–varying copula models with the FX analysis were reported in [19].

Change–point tests on the copula parameters can improve substantially the dependence modelling. In Chapter 5 we present a change–point analysis method for copulae based on the generalised likelihood ratio test. The distribution of the test is computed for some parametric copula families. We investigate the power of the test and the construction of confidence intervals. We illustrate the methods in a simulated example and apply them to an insurance data set. The analysis of change–points for the

copula of the FX residuals is also studied in this chapter. GARCH type models are fitted to the margins and then a copula change–point analysis of the filtered returns is performed. The change–points found are linked to macro–economic events. The material of Chapter 5 can be found in [18] and [19].

Finally, we summarise the methods presented and the results obtained.

Remark. The software developed for the various analyses is not presented in the thesis. Interested readers can contact the author for details (Email: `alexandra.dias@math.ethz.ch`).

Chapter 1

Inference for copulae

There are three main approaches to parameter inference for copulae. Being parametric models we can perform maximum likelihood estimation of the parameters and use asymptotic theory to derive standard errors and confidence intervals. As usual this consists of optimising a likelihood function in order to estimate the parameters. The optimisation is performed simultaneously in all the parameters and a simple bivariate model can already have five of them, two for each margin and one for the dependence structure. This number can easily grow leading to challenging numerical and computational problems. Using a two–step procedure we can decrease the complexity of the estimation procedure. First the margins are modelled and then the copula is fitted. The resulting inference procedure typically consists of a first parametric or non–parametric stage, followed by a parametric one. This approach is at least computationally less time consuming. In both cases it is possible to approximate standard errors and provide confidence intervals for the estimated model parameters.

The main definitions, properties and methods for copula–models inference are given is the next sections. Then, the methods are illustrated in an application to weekly FX returns on USD/DEM and USD/JPY spot rates.

1.1 Definitions, properties and examples

The main text book references for copulae are Joe [48] and Nelsen [63]. Further references are given throughout the text. We start with the definition of a copula.

Definition 1.1 *A copula is a multivariate distribution with standard uniform* $(0,1)$ *margins.*

An equivalent definition that provides some copula properties follows.

Definition 1.2 *A copula is a function* $C : [0,1]^d \to [0,1]$ *satisfying the conditions:*

i) For all $(u_1, u_2, \ldots, u_d)$ *in* $[0,1]^d$*, if at least one component* u_i *is zero, then* $C(u_1, u_2, \ldots, u_d) = 0$*;*

ii) For $u_i \in [0,1]$*,* $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ *for all* $i \in \{1, 2, \ldots, d\}$*;*

iii) For all $[u_{11}, u_{12}] \times [u_{21}, u_{22}] \times \ldots \times [u_{d1}, u_{d2}]$ *d–dimensional rectangles in* $[0,1]^d$*,*

$$\sum_{i_1=1}^{2} \ldots \sum_{i_d=1}^{2} (-1)^{i_1+\ldots+i_d} C(u_{1\,i_1}, u_{2\,i_2}, \ldots, u_{d\,i_d}) \geq 0.$$

The ability of a copula to separate the dependence structure from the marginal behaviour in a multivariate distribution comes from Sklar's theorem in [78].

Theorem 1.3 (Sklar's theorem)
Let F *be a* d*–dimensional distribution function with univariate margins* $F_1, F_2, \ldots, F_d$*. Let* R_i *be the range of* F_i*, for* $i = 1, 2, \ldots, d$*. Then there exists a unique function* H*, defined on* $R_1 \times R_2 \times \ldots \times R_d$ *such that*

$$F(x_1, x_2, \ldots, x_d) = H(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)).$$

The extension of the function H *to* $[0,1]^d$ *is a copula* C*. For* C *such an extension of* H*, we have that*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)).$$

Proof. See [48, page 41]. □

If $F_1, F_2, \ldots, F_d$ in Sklar's theorem are continuous, then the function H coincides with the copula C which is then unique.

Besides the ability of isolating the dependence structure of a multivariate distribution, the copula provides a way of constructing distributions from given margins. This result from [78] is given in the next theorem.

Theorem 1.4 (Sklar's theorem (continued))
Given univariate distribution functions $F_1, F_2, \ldots, F_d$ and a d–dimensional copula C, the function defined by

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \tag{1.1}$$

is a d–dimensional distribution function with univariate margins $F_1, F_2, \ldots, F_d$.

Proof. See [48, page 41]. □

So we are able to write the joint distribution as a function of the copula and the univariate marginal distributions. As the copula does not depend on the univariate marginal distributions, we can say that it captures completely all the information about the dependence between the variables. For this reason, from here on we will refer to a copula and a dependence structure indifferently. We will refer to a statistical model given by a multivariate distribution written like (1.1) as a copula–based model.

We will often make use of quantile functions, probability–integral and quantile transformations. Their definitions are as follows.

Definition 1.5 *Consider the random variables $X, X_1, X_2, \ldots, X_d$ with distribution functions $F, F_1, F_2, \ldots, F_d$ respectively.*

i) The quantile function of F is defined for all u in $(0, 1)$ by the generalised inverse of F:

$$F^{\leftarrow}(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\}.$$

ii) The probability–integral transformation is the mapping $T : \mathbb{R}^d \to [0,1]^d$, $(x_1, x_2, \ldots, x_d) \mapsto (F_1(x_1), F_2(x_2), \ldots, F_d(x_d))$. *The quantile transformation is the operation* $T^{\leftarrow} : [0,1]^d \to \mathbb{R}^d$, $(u_1, u_2, \ldots, u_d) \mapsto (F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2), \ldots, F_d^{\leftarrow}(u_d))$.

It is well–known that, if U is a uniform random variable on $(0,1)$, then $F^{\leftarrow}(U)$ has distribution function F. On the other hand, the distribution function F of a random variable X is continuous if and only if $F(X)$ is uniformly distributed on $(0,1)$.

Suppose that F_i is continuous and write $F_i(x_i) = u_i$ for $i = 1, 2, \ldots, d$. Making these substitutions in (1.1), we obtain

$$C(u_1, u_2, \ldots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)) \tag{1.2}$$

for $(u_1, u_2, \ldots, u_d) \in [0,1]^d$, which is an explicit expression for the copula as a function of the joint and the univariate marginal distribution functions.

REMARK. F^{-1} denotes the usual inverse function of F and is a particular case of $F^{\leftarrow}$ when the function F is continuous and strictly increasing. ▩

Example 1.6 (The logistic copula)
Consider the bivariate logistic distribution function given by

$$F(x,y) = \exp\left(-\left(x^{-\theta} + y^{-\theta}\right)^{1/\theta}\right), \quad x > 0,\ y > 0,\ \theta \geq 1.$$

We can rewrite this function in a more convenient form as follows:

$$F(x,y) = \exp\left(-\left(\left(-\log e^{-1/x}\right)^{\theta} + \left(-\log e^{-1/y}\right)^{\theta}\right)^{1/\theta}\right). \tag{1.3}$$

The margins of a bivariate logistic are standard Fréchet, i.e., they have distribution function $F(z) = e^{-1/z}$ for $z > 0$. So, we can readily obtain the copula C of a pair of random variables with bivariate logistic distribution from (1.3) using (1.2), leading to

$$C(u_1, u_2) = \exp\left(-\left((-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right)^{1/\theta}\right), \tag{1.4}$$

where $\theta \geq 1$ and $(u_1, u_2) \in [0,1]^2$.

The family of copulae in Example 1.6 is known as Gumbel, Gumbel–Hougaard or logistic. As we will see later, the characteristics of this family make it frequently suitable for modelling financial data.

The dependence structure between random variables does not change under increasing and continuous transformations of the margins. This invariance property is a very interesting feature of copulae and is formally stated in the next proposition.

Proposition 1.7 (Invariance property)
Let C be the copula of the random vector $(X_1, X_2, \ldots, X_d)$ and R_i the range of X_i for $i = 1, 2, \ldots, d$. If the functions $g_i : R_i \to \mathbb{R}$ for $i = 1, 2, \ldots, d$ are continuous and strictly increasing a.s. then C is still the copula of $(g_1(X_1), g_2(X_2), \ldots, g_d(X_d))$. If the univariate marginal distributions are continuous then the functions g_i only have to be increasing a.s. in order to keep the invariance of C.

Proof. See [29]. □

This invariance means that copulae are the natural framework to study dependence properties which are invariant under increasing transformations of the individual random variables. Suppose that we have a set of dependent insurance policies and we want to come up with a multivariate distribution as a model for the losses. The joint distribution for the losses will have the same copula as the one for the logarithm of the losses or for the integral–transforms of the losses. So, in terms of copula modelling, we are indifferent with which of the three marginal scales we work.

REMARK. Consider that X_1 and X_2 are random variables with continuous distribution functions F_1 and F_2 respectively. Note that $F_2^{\leftarrow} \circ F_1$ satisfies the conditions of Proposition 1.7. Applying this kind of transformation we can map each margin into a more convenient space for statistical fitting purposes or to better visualise the dependence structure, without changing the copula of the data. ▨

For some applications, the interesting model is sometimes defined via the so–called survival family rather than the copula family itself. The survival copula appears as the function which relates the joint survival function

of a multivariate distribution with the survival functions of the univariate margins.

Proposition 1.8 (Survival copula)
Let F be the distribution function of the random vector $(X_1, X_2, \ldots, X_d)$, with marginal distribution functions $F_1, F_2, \ldots, F_d$ respectively. Then, there exists a copula $\widehat{C}$ such that

$$\overline{F}(x_1, x_2, \ldots, x_d) = \widehat{C}(\overline{F}_1(x_1), \overline{F}_2(x_2), \ldots, \overline{F}_d(x_d)),$$

where $\overline{F}(x_1, x_2, \ldots, x_m) = P(X_1 > x_1, X_2 > x_2, \ldots, X_m > x_m)$ for $m = 1, \ldots, d$. Moreover, in the bivariate case if C is the copula of (X_1, X_2) then

$$\widehat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

Proof. See [63, page 28]. □

This section ends with some examples of copulae important for finance and insurance modelling. First we give an example of a so–called explicit copula and then two examples of implicit copulae. An implicit copula is obtained from a multivariate distribution function through (1.2). If the expression which defines the copula is not written as a function of the joint and marginal distribution functions the copula is said to be explicit.

Example 1.9 The Clayton copula with parameter $\theta \in (0, +\infty)$ is given by:

$$C(u_1, \ldots, u_d) = \left(\sum_{i=1}^{d} \left(u_i^{-\theta} - 1 \right) + 1 \right)^{-1/\theta}, \quad (u_1, \ldots, u_d) \in [0, 1]^d.$$

Example 1.10 Gaussian copula: Let $\mathbf{X}$ be a d–dimensional random vector. Suppose that $\mathbf{X}$ has a multivariate normal distribution with mean vector zero and correlation matrix Σ. The Gaussian copula is defined as

$$\begin{aligned} C(u_1, \ldots, u_d) &= P\left(\Phi_1(X_1) \leq u_1, \ldots, \Phi_1(X_d) \leq u_d\right) \\ &= P\left(X_1 \leq \Phi_1^{-1}(u_1), \ldots, X_d \leq \Phi_1^{-1}(u_d)\right) \\ &= \Phi_d\left(\Phi_1^{-1}(u_1), \ldots, \Phi_1^{-1}(u_d)\right), \end{aligned}$$

where $(u_1, \ldots, u_d) \in [0, 1]^d$ and Φ_i represents the distribution function of a standard normal i–dimensional random vector.

Example 1.11 t–copula: Suppose that $\mathbf{X}$ has a zero mean d–dimensional multivariate t–distribution with density function

$$f(x_1,\ldots,x_d) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^d\,|\Sigma|}}\left(1+\frac{\mathbf{x}^t\Sigma^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}}, \tag{1.5}$$

where Σ is the correlation matrix and ν are the degrees of freedom. Analogously to the Gaussian copula, the t–copula is given by

$$C(u_1,\ldots,u_d) = t_d\left(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_d)\right),$$

where $(u_1,\ldots,u_d)\in[0,1]^d$, t_d represents the distribution function of a d–dimensional random vector with density (1.5) and t_ν denotes the distribution function of a standard univariate student–t random variable with ν degrees of freedom.

1.2 Tail–dependence

An asymptotic measure of dependence specially focused on bivariate extreme values is the tail–dependence coefficient. For continuous marginal distributions the notion of tail–dependence coefficient is in fact a copula property and so we will follow here the definition in terms of copulae given by Joe [48].

Definition 1.12 *Let (U_1,U_2) be a vector of two random variables uniformly distributed on $(0,1)$ and such that*

$$\lim_{u\to 0^+} P(U_1\le u|U_2\le u) = \lambda_L$$

exists. If $\lambda_L\in(0,1]$ then C has lower tail–dependence and C has no lower tail–dependence if $\lambda_L = 0$. Similarly, if

$$\lim_{u\to 1^-} P(U_1> u|U_2> u) = \lambda_U$$

exists, C has upper tail–dependence if $\lambda_U\in(0,1]$ and has no upper tail–dependence if $\lambda_U = 0$.

Proposition 1.13 *Let C as a copula be the distribution function of the bivariate vector of standard uniform random variables (U_1, U_2). If λ_L exists then*

i) $\lambda_L = \lim_{u \to 0^+} C(u,u)/u$ *and*

ii) $\lambda_L = \lim_{u \to 0^+} (P(U_2 \leq u | U_1 = u) + P(U_1 \leq u | U_2 = u))$.

If λ_U exists then

iii) $\lambda_U = \lim_{u \to 0^+} \widehat{C}(u,u)/u$ *and*

iv) $\lambda_U = \lim_{u \to 1^-} (P(U_2 > u | U_1 = u) + P(U_1 > u | U_2 = u))$.

Proof. See [29] and [48]. □

We can easily see in Proposition 1.13 that if $C(u_1, u_2) = C(u_2, u_1)$, i.e. if C is exchangeable, then from *ii*)

$$\lambda_L = 2 \lim_{u \to 0^+} P(U_2 \leq u | U_1 = u)$$

and from *iv*) we have that

$$\lambda_U = 2 \lim_{u \to 1^-} P(U_2 > u | U_1 = u).$$

Often we will work with models where it makes sense to assume the continuity of the univariate marginal distributions. In this case, applying the quantile transformation to Definition 1.12 we obtain the next results.

Proposition 1.14 *If X_1 and X_2 have continuous distribution functions F_1 and F_2 respectively then*

$$\lambda_L = \lim_{u \to 0^+} P(X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u))$$

and

$$\lambda_U = \lim_{u \to 1^-} P(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u))$$

if the limits exist.

Proof. The results readily follow from applying the quantile transformation to the definitions of λ_L and λ_U. □

Whenever $\lambda_L \in (0,1]$ ($\lambda_U \in (0,1]$) we say that the random variables are asymptotically dependent in the lower (upper) tail. If $\lambda_L = 0$ ($\lambda_U = 0$) the random variables are said to be asymptotically independent in the lower (upper) tail.

Example 1.15 The tail–dependence coefficient for the t–copula from Example 1.11, takes the form

$$\lambda_L = \lambda_U = 2\bar{t}_{\nu+1}\left(\sqrt{(\nu+1)(1-\rho)/(1+\rho)}\right), \tag{1.6}$$

where $\bar{t}_\nu$ denotes the tail of a standard univariate t–distribution with ν degrees of freedom and ρ is the correlation parameter of the t–copula. This result is to be found in Embrechts et al. [29].

1.3 The IFM method

IFM stands for inference functions for margins. This terminology comes from McLeish and Small in [60] and it has been followed by authors like Joe [48] for copula–based model inference. The method consists of estimating the model parameters by finding the roots of a conveniently defined set of inference functions. In the case of maximum likelihood estimation the inference functions are the partial derivatives of the log–likelihood function. In the IFM method the score functions of the margins and of the copula constitute the set of estimating equations.

Consider the vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)^t$ of random variables. Suppose we want to estimate the parametric copula–based model for $\mathbf{X}$ given by

$$F(\mathbf{x}; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}) = C(F_1(x_1; \boldsymbol{\alpha}_1), F_2(x_2; \boldsymbol{\alpha}_2), \ldots, F_d(x_d; \boldsymbol{\alpha}_d); \boldsymbol{\theta}) \tag{1.7}$$

where $F_i(x_i; \boldsymbol{\alpha}_i)$ is the distribution function of X_i with parameter vector $\boldsymbol{\alpha}_i \in \mathbb{R}^{p_i}$ with $p_i \in \mathbb{N}$ for all univariate margins $i = 1, 2, \ldots, d$ and C is a copula family parameterised by the vector $\boldsymbol{\theta} \in \mathbb{R}^q$ with $q \in \mathbb{N}$. Assume that C has a density c given by

$$c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}) = \frac{\partial^d C(u_1, u_2 \ldots, u_d; \boldsymbol{\theta})}{\partial u_1 \, \partial u_2 \ldots \partial u_d}, \tag{1.8}$$

with $(u_1, u_2, \ldots, u_d) \in [0,1]^d$ and that F_i has a density f_i for all $i = 1, 2, \ldots, d$. For the case where the margins are discrete we denote by f_i the probability mass function of X_i. As for our applications margins are absolutely continuous, this will not be an issue. The density of the copula–model (1.7) for $\mathbf{X}$ is

$$f(\mathbf{x}; \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}) = c(F_1(x_1; \boldsymbol{\alpha}_1), \ldots, F_d(x_d; \boldsymbol{\alpha}_d); \boldsymbol{\theta}) \prod_{i=1}^{d} f_i(x_i; \boldsymbol{\alpha}_i). \tag{1.9}$$

Suppose that we have n iid d–dimensional vectors of observations $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. We assume that all the necessary regularity conditions (see [55]) on c and f_i for $i = 1, 2, \ldots, d$ are met. The log–likelihood function for the univariate margins of model (1.7) takes the form

$$L_i(\boldsymbol{\alpha}_i; \mathbf{x}) = \sum_{j=1}^{n} \log f_i(x_{ij}; \boldsymbol{\alpha}_i), \quad i = 1, 2, \ldots, d, \tag{1.10}$$

while the log–likelihood for the copula–model F is

$$L(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}; \mathbf{x}) = \sum_{j=1}^{n} \log f(\mathbf{x}_j; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}). \tag{1.11}$$

Supposing that the usual regularity conditions are fulfilled, there is a vector solution $\tilde{\boldsymbol{\alpha}}_i$ to each one of the d systems of equations

$$\left(\frac{\partial L_i(\boldsymbol{\alpha}_i; \mathbf{x})}{\partial \alpha_{1i}}, \frac{\partial L_i(\boldsymbol{\alpha}_i; \mathbf{x})}{\partial \alpha_{2i}}, \ldots, \frac{\partial L_i(\boldsymbol{\alpha}_i; \mathbf{x})}{\partial \alpha_{p_i i}} \right) = \mathbf{0}, \quad i = 1, 2, \ldots, d, \tag{1.12}$$

which is the maximum likelihood estimator (MLE) for the marginal parameters. Note that these estimators are obtained independently for each margin. The IFM method consists first of obtaining the MLE vectors $\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_d$ for the marginal parameters solving (1.12). Then substitute these marginal estimates in (1.11) to maximise (usually numerically)

$$L(\boldsymbol{\theta}; \mathbf{x}, \tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, \ldots, \tilde{\boldsymbol{\alpha}}_d) = \sum_{j=1}^{n} \log f(\mathbf{x}_j, \tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, \ldots, \tilde{\boldsymbol{\alpha}}_d; \boldsymbol{\theta}), \tag{1.13}$$

in order to estimate $\boldsymbol{\theta}$, or to use the score function of (1.11) and estimate the dependence parameter vector $\tilde{\boldsymbol{\theta}}$ solving the following system of

equations

$$\left(\frac{\partial L(\boldsymbol{\theta}; \mathbf{x}, \tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_d)}{\partial \theta_1}, \ldots, \frac{\partial L(\boldsymbol{\theta}; \mathbf{x}, \tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_d)}{\partial \theta_q}\right) = \mathbf{0}. \tag{1.14}$$

This procedure is far more easy numerically and less computationally intensive than a direct optimisation of (1.11). But, in case it is feasible to obtain the MLE for the full vector of model parameters from (1.11) we can use the IFM estimates as good starting values for the optimisation routine.

Variance–covariance matrix for IFM estimates

Theoretically, it is possible to derive the asymptotic variance–covariance matrix of $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_d, \tilde{\boldsymbol{\theta}})$, the IFM parameter estimators of a copula–based model like in (1.7). For completeness, we state this result below. For our applications this result is less useful as it leads to the computation of too many derivatives. Instead, we will use the jackknife method to compute an approximation to the asymptotic variance–covariance matrix.

The IFM estimates are the solutions of (1.12) and (1.14), the set of inference functions. Each one of these inference functions is the summand over all the observations of the derivative of a log–density function. Denote the terms of these summands by g, i.e. for the d margins we have the inference functions

$$\sum_{t=1}^{n} g_k(x_{ti}, \alpha_{ji}) = \sum_{t=1}^{n} \frac{\partial L_i(\boldsymbol{\alpha}_i; x_{ti})}{\partial \alpha_{ji}}$$

for $i = 1, 2, \ldots, d$, $j = 1, 2, \ldots, p_i$ and $k = p_1 + p_2 + \ldots + p_{i-1} + j$. Additionally, consider

$$\sum_{t=1}^{n} g_k(\mathbf{x}_t, \theta_k) = \sum_{t=1}^{n} \frac{\partial L(\boldsymbol{\eta}; \mathbf{x}_t)}{\partial \theta_k}$$

for $k = p + 1, p + 2, \ldots, p + q$ and $p = \sum_{i=1}^{d} p_i$. The estimators $\tilde{\boldsymbol{\eta}}$ satisfy the system of equations

$$\sum_{t=1}^{n} \mathbf{g}(\mathbf{x}_t, \tilde{\boldsymbol{\eta}}) = \mathbf{0}.$$

Let $\partial \mathbf{g}^t/\partial \boldsymbol{\eta}$ be a matrix with component (k,i) given by $\partial g_k(\mathbf{x},\boldsymbol{\eta})/\partial \eta_i$. The next result gives the asymptotic variance–covariance matrix of $\tilde{\boldsymbol{\eta}}$.

Theorem 1.16 *Suppose that $X_1, X_2, \ldots, X_d$ are iid with multivariate density given by (1.7) fulfilling the regularity conditions for asymptotic maximum likelihood theory. Let the vector of estimating equations for $\tilde{\boldsymbol{\eta}}$ be*

$$\sum_{t=1}^{n} \mathbf{g}(\mathbf{X}_t, \tilde{\boldsymbol{\eta}}) = \mathbf{0}.$$

The asymptotic variance–covariance matrix of $n^{1/2}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta})^t$, called the Godambe information matrix, is given by

$$V = D_{\mathbf{g}}^{-1} M_{\mathbf{g}} (D_{\mathbf{g}}^{-1})^t,$$

where

$$D_{\mathbf{g}} = E(\partial \mathbf{g}^t(\mathbf{X},\boldsymbol{\eta})/\partial \boldsymbol{\eta}), \quad M_{\mathbf{g}} = E(\mathbf{g}^t(\mathbf{X},\boldsymbol{\eta})\mathbf{g}(\mathbf{X},\boldsymbol{\eta})).$$

Proof. See [48, page 301]. □

In order to estimate the variance–covariance matrix of $\tilde{\boldsymbol{\eta}}$, which is $n^{-1}V$, we have to compute several derivatives for obtaining $E(\partial \mathbf{g}^t(\mathbf{X},\boldsymbol{\eta})/\partial \boldsymbol{\eta})$. These calculations can be very tedious and so it is very useful to have an alternative fast and reliable procedure. The jackknife method yields such a procedure.

Proposition 1.17 *Let $\tilde{\boldsymbol{\eta}}^{(-i)}$, for $i = 1, 2, \ldots, n$, be the IFM estimate of $\boldsymbol{\eta}$ obtained from the observed sample with the ith observation $\mathbf{x}_i$ deleted. The jackknife estimate of $n^{-1}V$ is*

$$\sum_{i=1}^{n} (\tilde{\boldsymbol{\eta}}^{(-i)} - \tilde{\boldsymbol{\eta}})(\tilde{\boldsymbol{\eta}}^{(-i)} - \tilde{\boldsymbol{\eta}})^t \tag{1.15}$$

where $\tilde{\boldsymbol{\eta}}^{(-i)}$ and $\tilde{\boldsymbol{\eta}}$ are considered as column vectors.

Proof. See [48, page 302]. □

The vector $\tilde{\boldsymbol{\eta}}^{(-i)}$ in Proposition 1.17 is the so–called leave–one–out jackknife estimate of $\boldsymbol{\eta}$. The jackknife approximation for the estimators variance is useful not only for the parameter estimates but also for other

estimators depending on model parameters. Proposition 1.17 can be generalised to real functionals of the model parameters. From Joe [48] we obtain that for a real–valued function b, the jackknife estimate of the standard errors of $b(\tilde{\boldsymbol{\eta}})$ is

$$\left(\sum_{i=1}^{n}\left(b\left(\tilde{\boldsymbol{\eta}}^{(-i)}\right)-b\left(\tilde{\boldsymbol{\eta}}\right)\right)^2\right)^{1/2}. \tag{1.16}$$

This is useful for evaluating the standard deviation of quantities that can be written as functions of model parameters. This is for instance the case for Spearmann's ρ and Kendall's τ for some copula families. For example, from the definition of Kendall's τ as a function of the copula (see [75]) we have that $\tau = 1 - 1/\theta$ for the Gumbel copula defined as in (1.4). In the case of the Clayton copula, given in Example 1.9, we have that $\tau = \theta/(\theta+2)$. For some more examples see [38]. In Section 1.7 we will use the jackknife method to estimate the standard errors not only for the parameters but also for the tail–dependence coefficient.

1.4 The pseudo log–likelihood method

In the IFM method for copula–based models we have to assume a parametric family for each of the margins. The choice of the "best" distribution for the margins is of course crucial. Hence, we need a procedure that avoids marginal model risk as much as possible and also does not mask the dependence structure too much. Several authors, see [13], [36], [46], [64], [65], have proposed in different fields of application a semi–parametric method tailor–made for copula–based models. Their proposal is the pseudo log–likelihood method. The estimation procedure consists of two stages. In the first one we model the margins using a non–parametric estimator and in the second step a parametric copula is fitted. Suppose that $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is a sample of n iid d–dimensional vectors. The difference for the IFM method is that here the distribution function of margin i is modelled by the rescaled empirical distribution function

$$F_{in}(x) = \frac{1}{n+1}\sum_{j=1}^{n}\mathbb{I}_{\{y\in\mathbb{R}\,:\,y\leq x\}}(x_{ij}),$$

assuming that the sample size will be large enough to enable a statistically accurate non–parametric estimation. As usual $\mathbb{I}_A$ denotes the indicator function of the set A. The n observed vectors $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{di})^t$ for $i = 1, 2, \ldots, n$ are then transformed to the so–called pseudo–observations

$$\left(F_{1n}\left(x_{1i}\right), F_{2n}\left(x_{2i}\right), \ldots, F_{dn}\left(x_{di}\right)\right), \tag{1.17}$$

for $i = 1, 2, \ldots, n$. Using these transformed observations, we obtain the estimates for the copula parameters maximising the pseudo log–likelihood function

$$L(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^{n} \log c(F_{1n}(x_{i1}), F_{2n}(x_{i2}), \ldots, F_{dn}(x_{id}); \boldsymbol{\theta}) \tag{1.18}$$

where c is the copula density of model (1.1) and is given by (1.8). Alternatively we can obtain the estimate $\hat{\boldsymbol{\theta}}$ solving the system

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log c(F_{1n}(x_{i1}), F_{2n}(x_{i2}), \ldots, F_{dn}(x_{id}); \boldsymbol{\theta}) = 0$$

for $j = 1, 2, \ldots, q$.

Consistency and asymptotic distribution

In Genest et al. [37] it is shown that the estimator $\hat{\boldsymbol{\theta}}$ for the dependence parameter resulting from (1.18) is consistent and has asymptotically normal distribution under regularity conditions similar to those of maximum likelihood theory. To recall here that result, consider the notation

$$\begin{aligned}
l(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d) &= \log c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}) \\
l_{\theta_i}(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d) &= \frac{\partial}{\partial \theta_i} l(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d), \quad i = 1, 2, \ldots, q \\
l_{\theta_i \theta_j}(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d) &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d), \quad i, j = 1, 2, \ldots, q \\
l_i(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d) &= \frac{\partial}{\partial u_i} l(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d), \quad i = 1, 2, \ldots, d \\
l_{\theta_i, j}(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d) &= \frac{\partial^2}{\partial \theta_i \partial u_j} l(\boldsymbol{\theta}; u_1, u_2, \ldots, u_d), \quad \begin{array}{l} i = 1, 2, \ldots, q, \\ j = 1, 2, \ldots, d. \end{array}
\end{aligned}$$

For the case of a bivariate model where the copula has a scalar parameter, $d = 2$ and $q = 1$, we have the following proposition.

Proposition 1.18 *In case θ is a scalar and $d = 2$, under suitable regularity conditions, the semi–parametric estimator $\hat{\theta}$ obtained maximising (1.18) is consistent and $n^{1/2}(\hat{\theta}-\theta)$ is asymptotically normal with variance $\nu^2 = \sigma^2/\beta^2$, where*

$$\beta = -E\left(l_{\theta,\theta}(\theta; F_1(X_1), F_2(X_2))\right) = E\left(l_\theta^2(\theta; F_1(X_1), F_2(X_2))\right)$$

and

$$\sigma^2 = var\left(l_\theta(\theta; F_{1n}(X_1), F_{2n}(X_2)) + W_1(X_1) + W_2(X_2)\right)$$

for $W_i(X_i) =$

$$\int_{[0,1]^2} \mathbb{I}_{\{u\in[0,1]\,:\,F_i(X_i)\leq u\}}(u_i) l_{\theta,i}(\theta; u_1, u_2) c(u_1, u_2; \theta) du_1 du_2 \qquad (1.19)$$

with $i = 1, 2$.

Proof. See [37]. □

An alternative expression for (1.19) which will be useful for estimation purposes is given, for $i = 1, 2$, by

$$W_i(X_i) = -\int_{[0,1]^2} \mathbb{I}_{\{u\in[0,1]\,:\,F_i(X_i)\leq u\}}(u_i) \cdot l_\theta(\theta; u_1, u_2) l_i(\theta; u_1, u_2) c(u_1, u_2; \theta) du_1 du_2.$$

With Proposition 1.18 we are equipped to construct confidence intervals and give standard errors for the estimate θ as long as we can estimate ν^2. Also in [37] such an estimator is provided.

An estimator for the variance of $\hat{\theta}$

As we could see previously, under the appropriate conditions the variance of the pseudo log–likelihood estimator is $\nu^2 = \sigma^2/\beta^2$ were β^2 can be viewed as the variance of the random variable

$$U(X_1, X_2) = l_\theta(\theta; F_1(X_1), F_2(X_2)),$$

given that the expected value of the score function $l_\theta(\theta; F_1(X_1), F_2(X_2))$ is zero, and σ^2 is the variance of

$$V(X_1, X_2) = U(X_1, X_2) + W_1(X_1) + W_2(X_2).$$

There exist estimators of those quantities, namely $\hat{\sigma}^2$ and $\hat{\beta}^2$, such that

$$\hat{\nu}^2 = \frac{\hat{\sigma}^2}{\hat{\beta}^2} \tag{1.20}$$

is a consistent estimator of ν^2. The variables U and V are not observed but we can compute the pseudo–observations $\hat{U}_i$ and $\hat{V}_i$ with $i = 1, 2, \ldots, n$. Estimates of σ^2 and β^2 are then obtained by the empirical sample variance estimator of the pseudo–observations. Suppose that $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ is a sample of n bivariate observations. The pseudo–observations $\hat{U}$ are given by

$$\hat{U}_i(x_{1i}, x_{2i}) = l_\theta(\hat{\theta}; F_{1n}(x_{1i}), F_{2n}(x_{2i})), \quad i = 1, 2, \ldots, n.$$

For the $\hat{V}$ pseudo–observations we need the rescaled empirical copula function of the sample,

$$C_n(u_1, u_2) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{I}_{\{(y_1,y_2)\in[0,1]^2 \,:\, y_1 \le u_1 \,,\, y_2 \le u_2\}}(u_{1i}, u_{2i}).$$

The pseudo–observations of $\hat{V}$ are then obtained from

$$\hat{V}_i(x_{1i}, x_{2i}) = \hat{U}_i(x_{1i}, x_{2i}) + \hat{W}_1(x_{1i}) + \hat{W}_2(x_{2i}), \quad i = 1, 2, \ldots, n,$$

with

$$\hat{W}_j(x_{ji}) = -\int_{[0,1]^2} \mathbb{I}_{\{(y_1,y_2)\in[0,1]^2 \,:\, F_{jn}(x_{ji}) \le y_j\}}(u_1, u_2) \cdot l_\theta(\hat{\theta}; u_1, u_2) l_j(\hat{\theta}; u_1, u_2) dC_n(u_1, u_2)$$

for $j = 1, 2$ and $i = 1, 2, \ldots, n$. Now suppose that we rearrange the sample $\{(x_{1i}, x_{2i}) : i = 1, 2, \ldots, n\}$ sorting the first components x_{1i} in increasing order. Denote the sample ordered in such a way by $\{(x_{1(i)}, x_{2(i)}) : i = 1, 2, \ldots, n\}$. As the pairs of observations must be kept coupled, applying the marginal empirical distributions, we obtain

$$\left\{(F_{1n}(x_{1(i)}), F_{2n}(x_{2(i)})) : i = 1, 2, \ldots, n\right\} = \\ = \left\{(i/(n+1), S_i/(n+1)) : i = 1, 2, \ldots, n\right\},$$

where S_i is the rank of x_{2i} within $\{x_{2i} : i = 1, 2, \ldots, n\}$. So we can estimate the pseudo–observations $\hat{U}$ as

$$\hat{U}_i(x_{1i}, x_{2i}) = l_\theta \left(\hat{\theta}; i/(1+n), S_i/(n+1)\right), \quad i = 1, 2, \ldots, n. \tag{1.21}$$

and $\hat{V}$ by

$$\hat{V}_i = \hat{U}_i - \frac{1}{n} \sum_{j=i}^{n} l_1 \left(\hat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_\theta \left(\hat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right)$$
$$- \frac{1}{n} \sum_{S_j \geq S_i} l_2 \left(\hat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_\theta \left(\hat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right)$$

for $i = 1, 2, \ldots, n$. The justification for the consistency of the estimator $\hat{\nu}^2$ is to be found in [37].

Efficiency of the pseudo log–likelihood estimator

As far as we know, the efficiency of the pseudo log–likelihood estimator is not yet completely clarified. Genest et al. [37] prove the estimator efficiency in case of the independence copula. For the Gaussian copula the same property is presented in [51]. But in both cases, the authors admit that these are exceptions rather than the norm within copula models. Genest and Werker [39] give a necessary and sufficient condition for the asymptotic efficiency of the pseudo log–likelihood estimator. They illustrate their result by showing that for the independence and for the Gaussian copula the efficiency condition is fulfilled, and that for the Farlie–Gumbel–Morgenstern family it is not; see [48] for the definition of this copula family. This result shows that a new semiparametric estimator for copula–based models where the margins are empirically modelled needs to be found. Some possible hints for research in this subject are proposed in [39]. Although we do not proceed in this direction in this thesis, for reasons of completeness, we give here a recent result concerning the efficiency of $\hat{\theta}$. Consider

$$I(\theta) = \int_{[0,1]^2} l_\theta^2(u_1, u_2) c(u_1, u_2; \theta) du_1 du_2$$

and, for $i = 1, 2$, within the family of primitives of

$$I_{\theta,i}(u_i) = \int_{[0,1]} l_\theta(u_1, u_2) l_i(u_1, u_2) c(u_1, u_2; \theta) du_{3-i}$$

let $W_i(u_i)$ denote the one which has zero expectation.

Proposition 1.19 *The pseudo log–likelihood estimator $\hat{\theta}$ is asymptotically efficient if and only if for each possible value of θ, there exists a constant $0 < I^*(\theta) \leq I(\theta)$ such that for all $(u_1, u_2) \in [0,1]^2$,*

$$\left(1 - \frac{I(\theta)}{I^*(\theta)}\right) l_\theta(u_1, u_2) = \sum_{i=1}^{2} E\left(W_i(U_i)|U_{3-i} = u_{3-i}\right) + \sum_{i=1}^{2} l_i(u_1, u_2) \int_{[0,u_{3-i}]} W_i(v) + E(W_{3-i}(U_{3-i})|U_i = v)dv.$$

Under these conditions, $I^*(\theta)$ is then the information for estimating θ in the semi–parametric model.

Proof. See [39]. □

The multiparameter case

Clearly, dependence in high–dimensional vectors can be very complicated and defies being appropriately modelled by copula models with a one–dimensional parameter. Already the very important t–copula model, even for $d = 2$, has two parameters. We therefore need to know the properties of pseudo log–likelihood estimates for the more–dimensional parameter case.

In [37] it is stated that a generalisation of Proposition 1.18 is possible yielding that the pseudo log–likelihood estimator for the vector parameter $\boldsymbol{\theta} \in \mathbb{R}^q$ is consistent and asymptotically normal. Moreover, the asymptotic variance–covariance matrix of $n^{1/2}\boldsymbol{\theta}$ is

$$B^{-1}\Sigma B^{-1} \tag{1.22}$$

where B is the information matrix associated with the copula and Σ is the variance–covariance matrix of the q–dimensional random vector with kth component

$$\frac{\partial}{\partial \theta_k} \log c(F_1(X_1), F_2(X_2), \ldots, F_d(X_d); \boldsymbol{\theta}) + \sum_{i=1}^{d} W_{\theta_k i}(X_i), \tag{1.23}$$

with $W_{\theta_k i}(X_i) =$

$$\int_{[0,1]^d} \mathbb{I}_{\{u\in[0,1]\,:\,F_i(X_i)\leq u\}}(u_i) l_{\theta_k i}(\boldsymbol{\theta}; u_1, \ldots, u_d) dC(u_1, \ldots, u_d; \boldsymbol{\theta}).$$

In order to clarify the generalisation of the variance estimator, suppose now that $\boldsymbol{\theta}$ has dimension q and $d = 2$. Given the properties for $q = 1$ we have as an estimator for the matrix B the following. The diagonal elements of B, B_{kk} can be estimated as the variance of the pseudo–observations $\hat{U}_{\theta_k i}$,

$$\hat{U}_{\theta_k i}(x_{1i}, x_{2i}) = l_{\theta_k}\left(\hat{\theta}_k; \frac{i}{1+n}, \frac{S_i}{1+n}\right),$$

for $k = 1, 2, \ldots, q$ and $i = 1, 2, \ldots, n$ and assuming that we ordered the sample just like for (1.21). The off–diagonal elements of B, B_{kj} for $k \neq j$ can be estimated by

$$\hat{B}_{kj} = -\frac{1}{n}\sum_{i=1}^{n} l_{\theta_k\theta_j}\left(\hat{\theta}_k, \hat{\theta}_j; \frac{i}{1+n}, \frac{S_i}{1+n}\right).$$

The matrix Σ can be estimated as the sample variance–covariance matrix of the q vectors of pseudo–observations $\hat{V}_{\theta_k i}$ for $k = 1, 2, \ldots, q$ and $i = 1, 2, \ldots, n$, namely

$$\begin{aligned}\hat{V}_{\theta_k i} = \hat{U}_{\theta_k i} &- \frac{1}{n}\sum_{j=i}^{n} l_1\left(\hat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_{\theta_k}\left(\hat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right) \\ &- \frac{1}{n}\sum_{S_j\geq S_i} l_2\left(\hat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_{\theta_k}\left(\hat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right).\end{aligned}$$

The generalisation to the multidimensional case where $d > 2$ follows in a straightforward way from the bivariate case.

1.5 Pseudo log–likelihood for dependent sequences

The serial or temporal iid assumption is often realistic in the case of insurance portfolios with dependent risks but seldom acceptable for financial

time series. In a time dependence context, the copula–based model fitted to the observations with the pseudo log–likelihood method leads to inference about the stationary contemporaneous dependence between the d time series.

Consider the time series of d–dimensional random vectors $(\mathbf{X}_t)_{t=1,\ldots,n}$ with stationary multivariate distribution function given by

$$F(\mathbf{x}; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}) = C(F_1(x_1; \boldsymbol{\alpha}_1), F_2(x_2; \boldsymbol{\alpha}_2), \ldots, F_d(x_d; \boldsymbol{\alpha}_d); \boldsymbol{\theta}),$$

as in (1.7). Let $F_n(\mathbf{x})$ be the rescaled multivariate empirical distribution function of $F(\mathbf{x})$,

$$F_n(x_1, \ldots, x_d) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{I}_{\{(y_1,\ldots,y_d) \in \mathbb{R}^d \,:\, y_1 \leq x_1, \ldots, y_d \leq x_d\}}(x_{1i}, \ldots, x_{di}).$$

It is known, see Chen and Fan [12], Kiefer and Vogelsang [50] and references therein, that under mild conditions,

$$\sup_{x_i} |F_{in}(x_i) - F_i(x_i)| = o_{a.s.}(1), \quad \sqrt{n} \sup_{x_i} |F_{in}(x_i) - F_i(x_i)| = O_p(1)$$

for $i = 1, 2, \ldots, d$, as well as that

$$\sup_{\mathbf{x}} |F_n(\mathbf{x}) - F(\mathbf{x})| = o_{a.s.}(1), \quad \sqrt{n} \sup_{\mathbf{x}} |F_n(\mathbf{x}) - F(\mathbf{x})| = O_p(1).$$

In close spirit to the pseudo log–likelihood estimator from Genest et al. [37] and for appropriate conditions on the time series $(\mathbf{X}_t)_{t=1,\ldots,n}$, it seems natural to explore the maximiser $\hat{\boldsymbol{\theta}}$ of (1.18) as an estimator for the stationary dependence parameters. This estimator is not only the natural continuation of the iid case but it is also used by Chen and Fan [12] to estimate the copula parameter in a copula–based first–order Markov time series. See [48, Chapter 8] for an introduction to copula–based Markov time series and [16] for further details.

The pseudo log–likelihood method can also be used with time dependent multivariate observations to model not the stationary but the conditional copula. We fit copula models to the residuals obtained after having fitted univariate time series models to the margins. If each univariate model performs well, the residuals form a sequence of multivariate vectors independent in time and the methodology of Section 1.4 is applicable.

1.6 Goodness–of–fit test

A general goodness–of–fit test for a given copula family can be performed using the following well known result, the probability–integral transformation; see for instance Rosenblatt [73].

Let $\mathbf{X} = (X_1, X_2, \ldots, X_d)^t$ be a random vector with absolutely continuous distribution function $F(x_1, x_2, \ldots, x_d)$. Let $F_i(x_i) = P(X_i \leq x_i)$ be the distribution function of the univariate margins X_i, for $i = 1, \ldots, d$ and $F_{i|1,\ldots,i-1}(x_i|x_1, \ldots, x_{i-1}) = P(X_i \leq x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$, for $i = 2, 3, \ldots, d$. Consider the d transformations

$$\begin{aligned} T(x_1) &= P(X_1 \leq x_1) = F_1(x_1), \\ T(x_2) &= P(X_2 \leq x_2 | X_1 = x_1) = F_{2|1}(x_2|x_1), \\ &\vdots \\ T(x_d) &= P(X_d \leq x_d | X_1 = x_1, \ldots, X_{d-1} = x_{d-1}) \\ &= F_{d|1,\ldots,d-1}(x_d|x_1, \ldots, x_{d-1}). \end{aligned}$$

Then the random variables $Z_i = T(X_i)$, for $i = 1, \ldots, d$ are uniformly and independently distributed on $[0, 1]^d$.

Suppose now that C is a copula such that

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

If $C_i(u_1, \ldots, u_i)$ denotes the joint i–marginal distribution

$$C_i(u_1, \ldots, u_i) = C(u_1, \ldots, u_i, 1, \ldots, 1), \qquad i = 2, \ldots, d-1,$$

of $(U_1, \ldots, U_i)$, with $C_1(u_1) = u_1$ and $C_d(u_1, \ldots, u_d) = C(u_1, \ldots, u_d)$, then the conditional distribution of U_i, given the values of $U_1, \ldots, U_{i-1}$, is

$$C_i(u_i|u_1, \ldots, u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, \ldots, u_i)}{\partial u_1 \ldots \partial u_{i-1}} \Bigg/ \frac{\partial^{i-1} C_{i-1}(u_1, \ldots, u_{i-1})}{\partial u_1 \ldots \partial u_{i-1}},$$

for $i = 2, \ldots, d$. Hence we can write the variables Z_i, for $i = 2, \ldots, d$, using the conditional distributions C_i,

$$Z_i = C_i(F_i(X_i)|F_1(X_1), \ldots, F_{i-1}(X_{i-1})).$$

If $(F_1(X_1), F_2(X_2), \ldots, F_d(X_d))$ has distribution function C, then $\Phi^{-1}(Z_i)$, $i = 1, \ldots, d$, are iid, $\mathrm{N}(0,1)$ distributed and $S = \sum_{i=1}^{d} \left(\Phi^{-1}(Z_i)\right)^2$ has a chi–square distribution with d degrees of freedom. In particular for $d = 2$,

$$S = \left(\Phi^{-1}(F_1(X_1))\right)^2 + \left(\Phi^{-1}(C_2(F_2(X_2)|F_1(X_1)))\right)^2. \tag{1.24}$$

1.7 The IFM method for FX weekly returns

The aim of this section is threefold. First, we want to give an application of the IFM method. Second, we start our study of the statistical properties of multivariate FX time series. Finally, we illustrate how fitted copula–based models can be used to analyse financial data. For this preliminary analysis, we use weekly log–returns on the FX data. At such a low frequency, financial return observations are much closer to be iid than at the higher frequencies to be analysed later. Hence, weekly log–returns raise less time dependence and seasonality questions.

Figure 1.1 plots in two panels the $n = 649$ observed logarithmic weekly returns respectively on USD/DEM and USD/JPY spot rates from April 28, 1986 up to October 4, 1998. The series look fairly stationary and actually running the KPSS test for stationarity (see [53]) we get the test statistic values of 0.1382 for USD/DEM and 0.2802 for USD/JPY. Given that, for example the 90% quantile for the KPSS statistic is 0.349 and that it is a one–sided right–tailed test, we do not reject the stationarity hypothesis at this confidence level for neither of the series. For a simulation algorithm for the distribution of the KPSS test statistic see Zivot and Wang [82, page 124]. As the FX seasonalities are intra–day and intra–week, see Section 2.1, they are not present in weekly returns. On the other hand, the time dependence is also much weaker for that time horizon. We can inspect these two facts looking at the corresponding correlograms plotted in Figure 1.2. The plotted autocorrelograms, top left and bottom right panels in Figure 1.2, do not show evidence against time–independence in weekly returns. For the cross–correlograms the same observation can be made. Seasonalities are also not visible in the autocorrelograms. Nevertheless, we must say that both time series plotted in Figure 1.1 reveal the so called ARCH effect or conditional heteroscedasticity, a discrete–time version of stochastic volatility. This can be checked graphically by auto

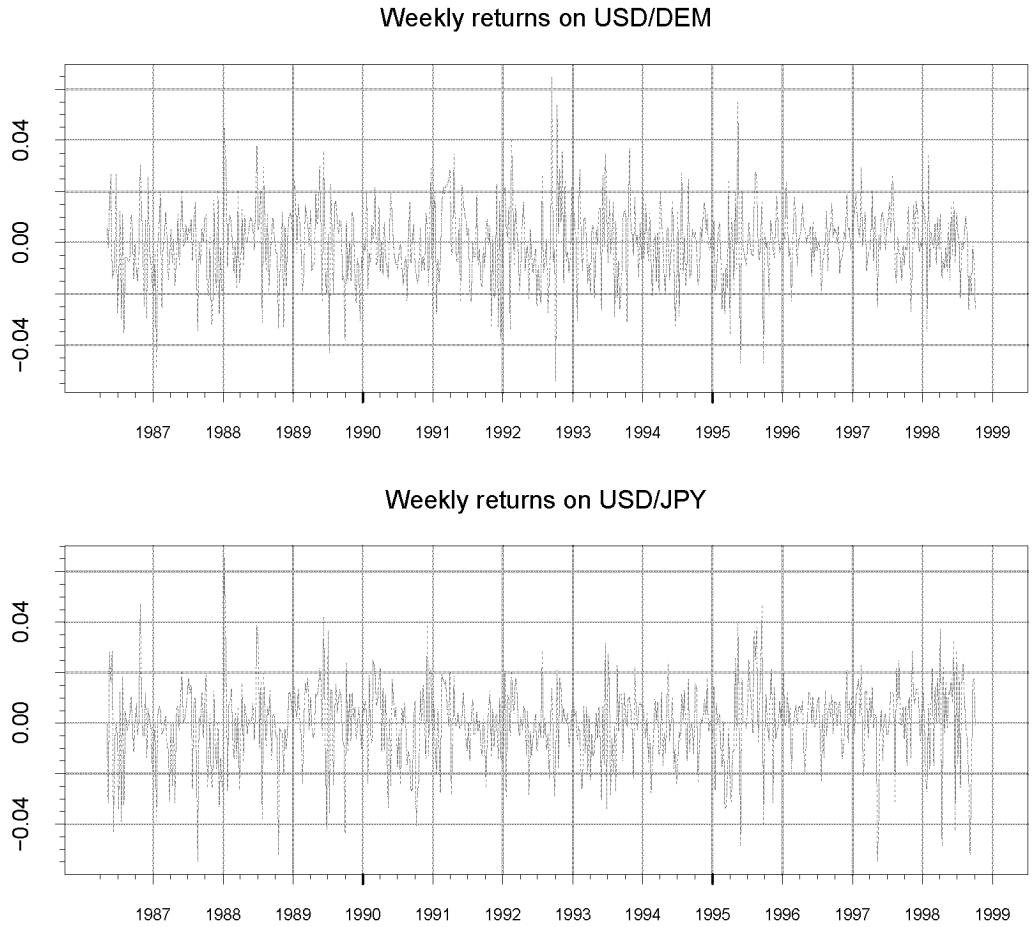

Figure 1.1: *Logarithmic weekly returns on USD/DEM (top) and USD/JPY (bottom) spot rates between April 28, 1986 and October 4, 1998.*

and cross–correlograms on the squared returns, not plotted here. We can also formally test for this effect on a time series (see Section 3.1 and [31]). The test consists of first fitting an ARCH model to the data. Then we test the null hypothesis that the squared residuals are not coming from an autoregressive process. If we reject the null hypothesis we conclude that the data have conditional heteroscedasticity. Performing this test we obtain zero p–values, up to the third decimal place, for both weekly return time series. This test procedure is presented in the first section of Chapter 3. We refrain from giving more details about this here as conditional heteroscedasticity will be considered in Chapters 3, 4 and 5. In this

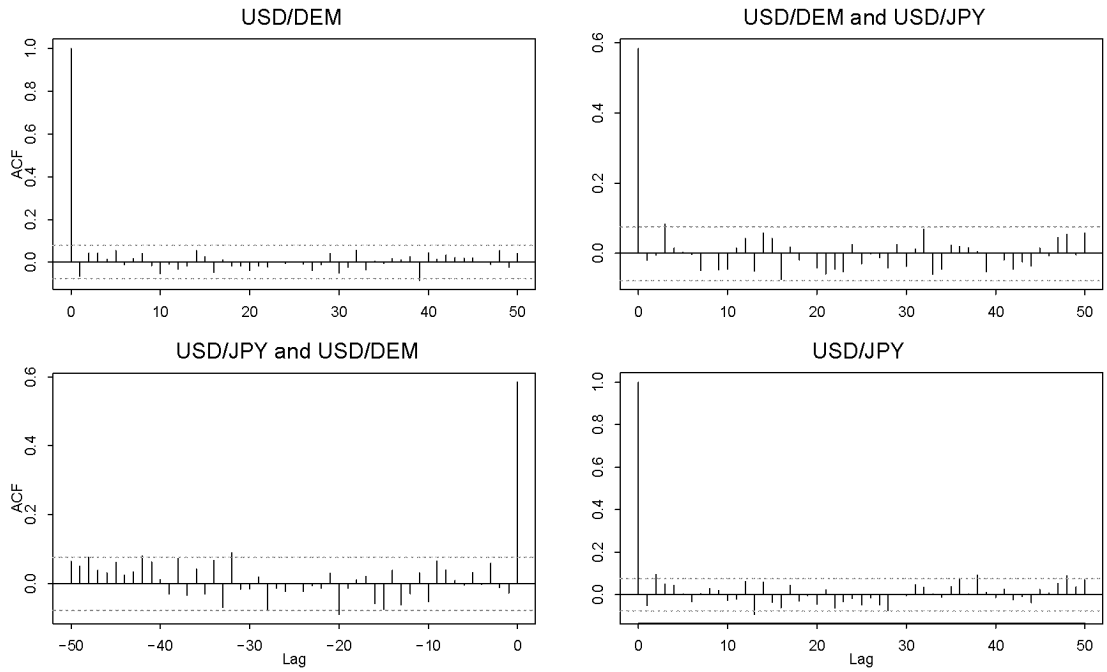

Figure 1.2: *Sample autocorrelation functions for logarithmic weekly returns of USD/DEM on the top left and USD/JPY on the bottom right. The sample cross–correlations are plotted for USD/DEM on past USD/JPY on the top right and USD/JPY on past USD/DEM on the bottom left.*

section we fit copula–based models using the IFM method to the weekly bivariate returns on USD/DEM and USD/JPY assuming these to be iid.

1.7.1 IFM estimates

In order to perform the first step in the IFM method, independently of any copula model, we have to choose a family of distributions to model each univariate return time series. To have an idea about the desirable shape of such families we plotted in Figure 1.3 the histograms for the returns. They look quite symmetric, especially the USD/DEM returns. From several alternative univariate models the best fitting was achieved in both cases with the univariate t–distribution with three parameters and density

$$f(x;\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)\sigma}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}, \qquad (1.25)$$

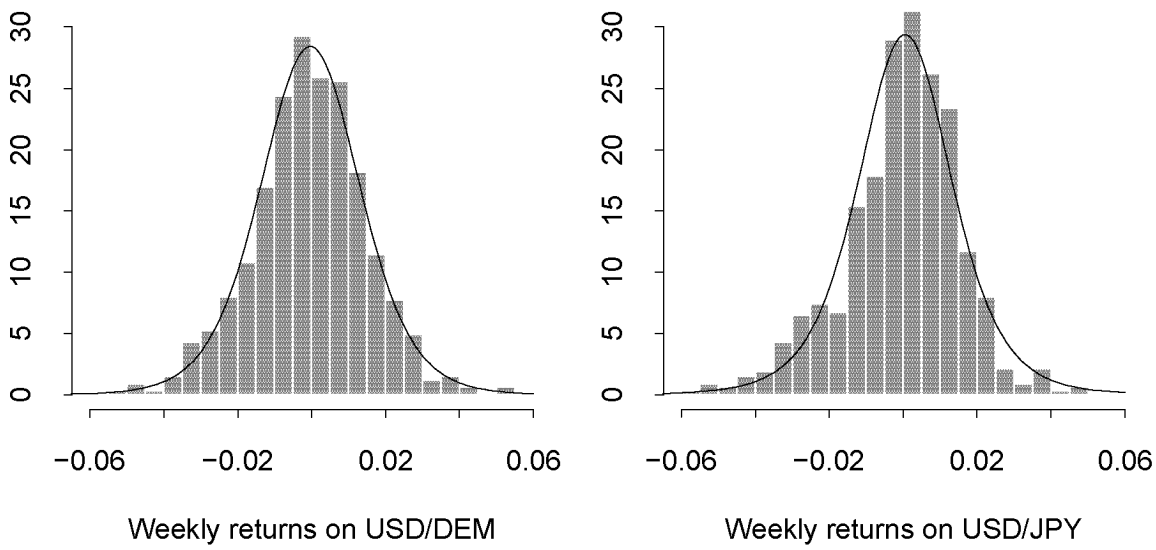

Figure 1.3: *Histograms of weekly returns of USD/DEM on the left and USD/JPY on the right. Super–imposed on the frequency bars are drawn the fitted densities to the data, a univariate t–distribution with three parameters.*

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\nu > 0$. The choice of a best fitting model is based on two factors. First, the models are ranked by their Akaike information criterion (AIC) given by

$$\text{AIC} = -2L(\alpha_1, \alpha_2, \ldots, \alpha_q; \mathbf{x}) + 2q, \tag{1.26}$$

where q is the number of parameters of the family of distributions fitted. The smaller the Akaike information value the better the model fits the data. Secondly, the Anderson–Darling [1] goodness of fit test is performed. We list in Table 1.1 the parameter estimates with the jackknife approximations (1.15) of the standard errors ($\hat{se}$.) and the Anderson–Darling p–values for the two marginal models.

Although we have good p–values for both models, we can inspect the quality of the fittings graphically through the quantile plots (QQ–plots),

$$\left\{ \left(x_{(k)}, F^{\leftarrow}\left(\frac{n-k+0.5}{n} \right) \right) : k = 1, 2, \ldots, n \right\},$$

where n is the number of observations, $x_{(k)}$ is the kth largest one and $F^{\leftarrow}$

Returns	$\hat{\mu}$ $(\hat{s.e.})$	$\hat{\sigma}$ $(\hat{s.e.})$	$\hat{\nu}$ $(\hat{s.e.})$	p–value
USD/DEM	-0.000367 (0.000592)	0.0136 (0.0007)	8.73 (3.06)	0.987
USD/JPY	0.000566 (0.000646)	0.0129 (0.0008)	5.42 (1.26)	0.210

Table 1.1: *Parameter estimates, $\hat{s.e.}$'s and p–values obtained from fitting univariate t–distributions to the weekly USD/DEM and USD/JPY returns.*

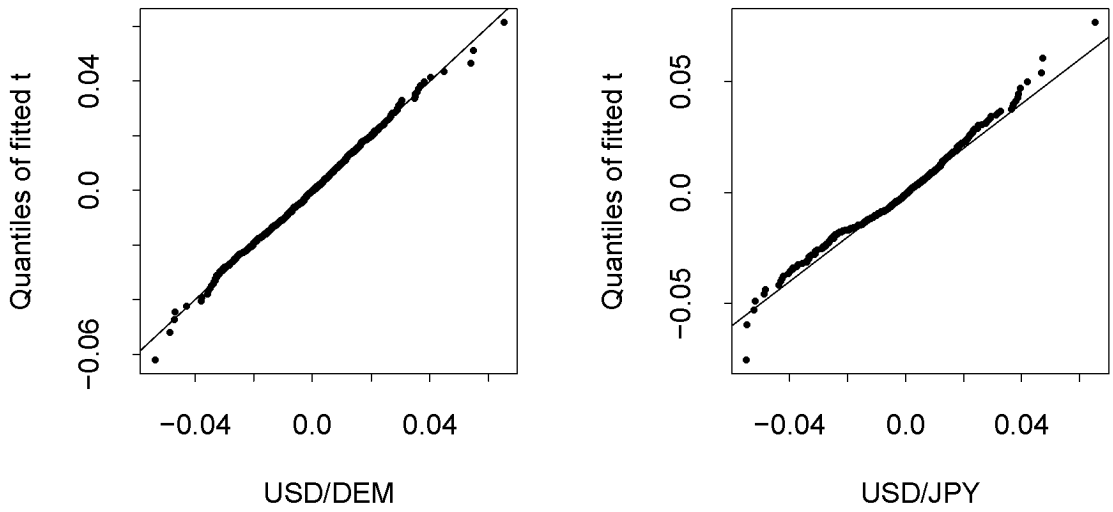

Figure 1.4: *QQ–plots of the data against the quantiles of the fitted t–distribution model for the two return series USD/DEM and USD/JPY.*

is the quantile function of a random variable with density (1.25). These are displayed in Figure 1.4. For the USD/DEM returns the QQ–plot is quite linear indicating a good fit and for the USD/JPY is less linear but still reasonable. These results agree with the goodness of fit p–values obtained where the USD/DEM has a much higher value than the USD/JPY. Of course, other models can be (and have been) tried. A versatile class widely used in finance are the so–called generalised hyperbolic distributions; see for instance Eberlein [21]. From various studies we know however that the t–distribution is a model acceptable on several economic grounds; see Platen [69].

Once we have a model for each of the individual series we can map the observations into the unit square applying the probability–integral transformation and proceed to the second stage of the IFM method fitting

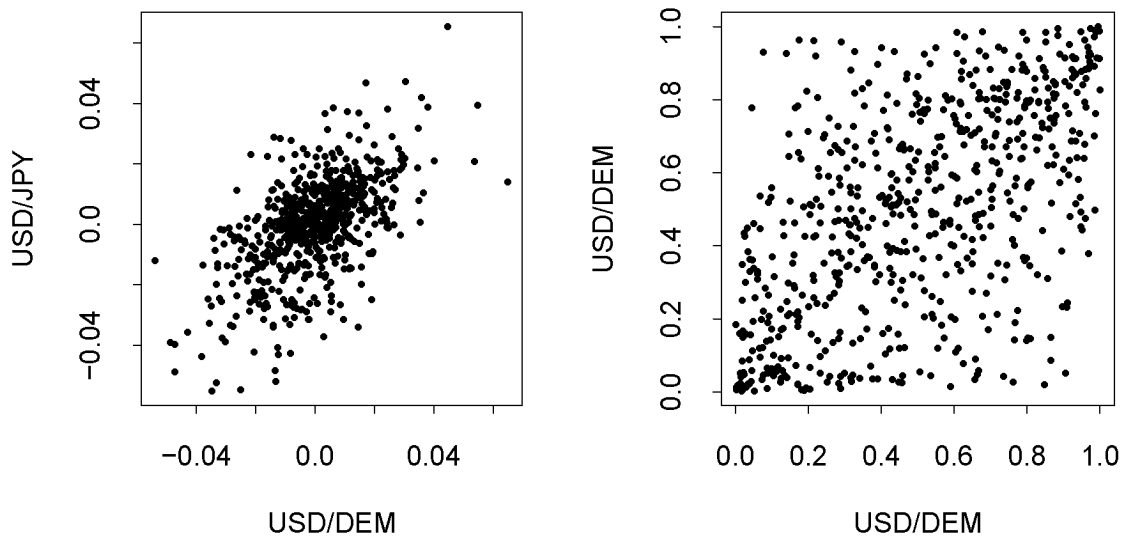

Figure 1.5: *Scatter–plots of weekly returns on USD/DEM and USD/JPY spot rates observed (in left panel) and mapped on* $[0,1]^2$ *by the fitted marginal models, the pseudo–observations (in the right panel).*

copula families to the pseudo–observations in (1.27).

In Figure 1.5 we have plotted the bivariate observed returns $\{(x_{1i}, x_{2i}) : i = 1, 2, \ldots, n\}$ on the left panel and on the right panel the probability transformed returns (pseudo–observations)

$$\left\{ \left(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}) \right) : i = 1, 2, \ldots, n \right\}, \tag{1.27}$$

where $\hat{F}_1$ and $\hat{F}_2$ are the fitted t–distributions to the USD/DEM and USD/JPY spot rate returns respectively. The most evident features in these plots are that the returns are positively dependent and fairly symmetric around the main diagonal. This second feature we see more clearly in the contour–plot of the pseudo–observations plotted in Figure 1.6. Useful exploratory indicators for the degree of dependence are the Pearson correlation coefficient r, Spearmann's ρ and Kendall's τ. The estimates obtained for these measures and the approximated standard errors $(s\hat{.}e.)$ are

$\hat{r}\,(s\hat{.}e.)$	$\hat{\rho}\,(s\hat{.}e.)$	$\hat{\tau}\,(s\hat{.}e.)$
0.5678 (0.0291)	0.5678 (0.0291)	0.4045 (0.0228)

These values were calculated on the pseudo–observations. Pearson's correlation coefficient, or linear correlation, depends on the univariate marginal

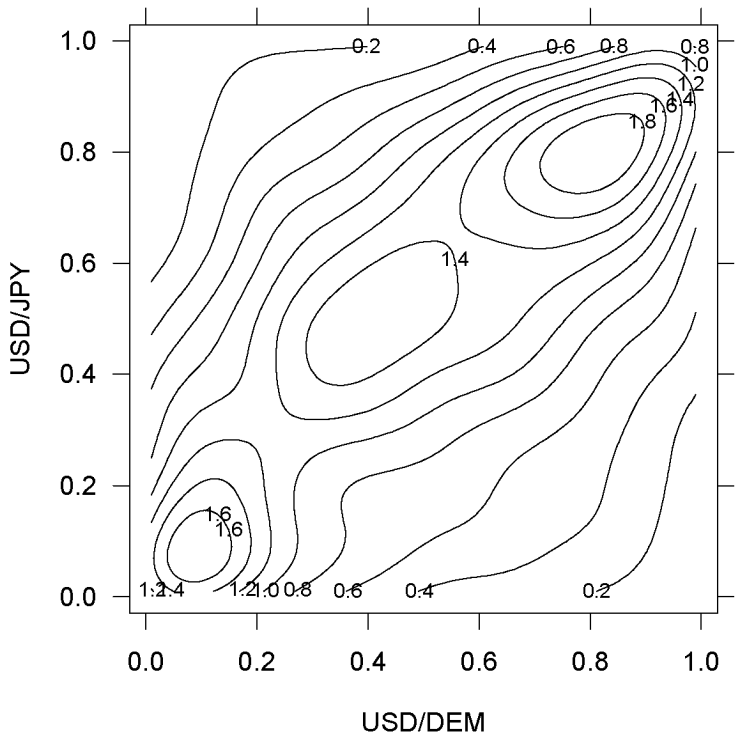

Figure 1.6: *Contour–plot of weekly returns on USD/DEM and USD/JPY spot rates mapped into* $[0,1]^2$ *by the fitted marginal models.*

distributions. This means that if we evaluate the dependence using linear correlation we may obtain different values for joint distributions with the same dependence structure and different margins. On the other hand, rank correlation measures like Spearman's ρ and Kendall's τ are independent of the univariate marginal distributions. For the properties of linear correlation, Spearman's ρ and Kendall's τ as measures of dependence, see Embrechts et al. [29] and references therein. Actually, the estimated Pearson correlation coefficient on the observed weekly FX returns is $\hat{r}\,(\hat{s.e.}) = 0.5834\,(0.0283)$.

Based on these observations and on the usual stylised facts for financial data, like tail–dependence for example, we will fit the following copula families to the data: t, Frank, Plackett, Gaussian, Gumbel, Clayton and the mixtures Gumbel with survival Gumbel, Clayton with survival Clayton, Gumbel with Clayton and survival Gumbel with survival Clayton. For further details on these copula families see [29], [48] and [63]. De-

Copula model	$\hat{\theta}$ (s.e.)	AIC	p–value
Clayton	0.812 (0.082)	-207.47	0.003
Frank	4.30 (0.32)	-255.24	0.574
Gaussian	0.584 (0.028)	-268.60	0.081
Gumbel	1.64 (0.06)	-258.40	0.330
Plackett	6.82 (0.78)	-264.80	0.944

Table 1.2: *Weekly returns on USD/DEM and USD/JPY. Estimates and standard errors of dependence parameters in Clayton, Frank, Gaussian, Gumbel and Plackett models. For each model fitted we provide the AIC and the p–value.*

noting the copula family A with parameter $\boldsymbol{\theta}$ by $C^A(\cdot,\cdot;\boldsymbol{\theta})$, the fitted mixtures have distribution functions of the form

$$C(u_1,u_2;\boldsymbol{\theta}) = \theta_3\, C^A(u_1,u_2;\theta_1) + (1-\theta_3)\, C^B(u_1,u_2;\theta_2).$$

We fitted all the listed models to the pseudo–observations (1.27) to obtain the parameter estimates, maximising the function

$$L(\boldsymbol{\theta};\mathbf{x},\ \tilde{\boldsymbol{\alpha}}_1,\tilde{\boldsymbol{\alpha}}_2) = \sum_{i=1}^{n} \log c(\hat{F}_1(x_{1i}),\hat{F}_2(x_{2i});\boldsymbol{\theta}) \tag{1.28}$$

which is equivalent to maximising (1.13). From the maximised log–likelihood we computed the AIC values for each family. We estimated the standard errors using the jackknife approximation (1.16) from Section 1.3. These values are all given in Tables 1.2 and 1.3. For the t–copula the parameters θ_1 and θ_2 in Table 1.3 represent respectively the degrees of freedom and the correlation. The p–values for each copula–based model given in the tables are computed using the statistic (1.24).

From the models fitted to the pseudo–observations, the one which has the best AIC is the mixture of 0.57 of Gumbel with 0.43 of survival Gumbel. According to the AIC criterion, the fitted models can be ordered within two groups as follows: Gumbel with survival Gumbel, t, survival Clayton with survival Gumbel, Clayton with Gumbel, Clayton with survival Clayton. In the second group we have by increasing AIC (best to worst model): Plackett, Gaussian, Gumbel, Frank and Clayton.

Copula model	$\hat{\theta}_1$ (s.e.)	$\hat{\theta}_2$ (s.e.)	$\hat{\theta}_3$(s.e.)	AIC	p–value
Clayton & surv. Clayton	1.06 (0.31)	1.53 (0.43)	0.492 (0.091)	-267.36	0.217
Clayton & Gumbel	1.69 (0.23)	1.19 (0.97)	0.703 (0.126)	-271.41	0.274
surv. Clayton & surv. Gumbel	1.55 (0.11)	1.84 (0.61)	0.635 (0.088)	-271.85	0.343
Gumbel & surv. Gumbel	1.76 (0.28)	1.55 (0.26)	0.565 (0.123)	-273.09	0.316
t	9.74 (4.86)	0.58 (0.02)	–	-272.65	0.216

Table 1.3: *Weekly returns on USD/DEM and USD/JPY. Estimates and standard errors of parameters for the t–model and for the four mixture models considered. In case of the mixture models, θ_1 and θ_2 are the dependence parameters respectively for the first and second terms of the mixture. θ_3 is the mixture parameter which gives the proportion of the first term. For the t–model, θ_1 are the degrees of freedom and θ_2 is the correlation. For each model fitted we provide the AIC and the p–value.*

The mixture models and the t–model perform better than the one–parameter models. The number of parameters can not be the main reason for this difference as the AIC is penalised by the number of model parameters. The mixture models perform similarly. Both copulae with very asymmetric tails give poor fits. We should remark that, although with a middle AIC rank, the Plackett copula has a very high goodness–of–fit p–value. It is not surprising that the Gaussian and the t–copula have significantly different AIC values. The limit of the t–copula when the degrees of freedom go to infinity is the Gaussian copula. The estimated degrees of freedom for the t–copula are too small to suggest that it may be close to the Gaussian.

From this analysis we can conclude that the weekly returns on USD/DEM and USD/JPY spot rates can be modelled by a mixture of the Gumbel and survival Gumbel copulae or by the t–copula, together with univariate t–distributions for the margins. With such a model we are equipped to estimate functionals of those returns like portfolio expected value, Value–at–Risk or expected shortfall risk measures. We just have to simulate from the distribution of the model selected and then estimate the functionals of interest.

1.7.2 VaR and expected shortfall estimation

In this section we illustrate how the fitted copula–based models can be used to estimate quantities of interest for risk management like Value–at–Risk (VaR) and expected shortfall (ES). These risk measures have the following probabilistic definitions.

Definition 1.20 *Consider a portfolio of risky assets, a fixed time horizon Δt and a confidence level $\alpha \in (0,1)$. Let L be the random variable representing the loss function for the portfolio and t the reference time.*

i) The portfolio Value–at–Risk at the confidence level α for the time horizon Δt, $VaR_{\alpha,t+\Delta t}$ is the α–quantile of the loss distribution, i.e.

$$VaR_{\alpha,t+\Delta t} = \inf\{l \in \mathbb{R} : P(L \leq l) \geq \alpha\} = F^{\leftarrow}(\alpha)$$

ii) The portfolio expected shortfall at the confidence level α for the time horizon Δt, $ES_{\alpha,t+\Delta t}$ is the conditional expectation

$$ES_{\alpha,t+\Delta t} = E(L|L \geq VaR_{\alpha,t+\Delta t}),$$

assuming of course that $E(L) < \infty$.

The level α typically takes values in $\{0.90, 0.95, 0.99\}$. We will drop the confidence level and the time subscripts in $\text{VaR}_{\alpha,t+\Delta t}$ and in $\text{ES}_{\alpha,t+\Delta t}$ whenever confusion can not arise. The portfolio loss distribution will be here the portfolio return with positive losses and negative gains.

Consider a portfolio composed of 50% of USD/DEM and 50% of USD/JPY at reference time t = 4 October 1998. From the data we have that, $P_{1,t} = 0.4953665$ and $P_{2,t} = 4.911505$ are respectively the spot rates of USD/DEM and USD/JPY on October 4, 1998. Additionally denote the portfolio weights of USD/DEM and USD/JPY respectively by $w_{1,t} = 0.5$ and $w_{2,t} = 0.5$ at the moment t. Assuming that we keep the weights of each asset constant during the next period, the portfolio value at the future time $t+1$ = 11 October 1998 will be given by the random variable

$$V_{t+1} = 0.5\, P_{1,t} \exp(X_{1,t+1}) + 0.5\, P_{2,t} \exp(X_{2,t+1}). \tag{1.29}$$

The weekly portfolio return at the same date reads

$$Y_{t+1} = \log\left(\frac{V_{t+1}}{v_t}\right), \tag{1.30}$$

where $X_{1,t+1}$ is the unknown weekly return of USD/DEM at $t+1$, $X_{2,t+1}$ is the weekly return of USD/JPY at the same time and v_t is the observed portfolio value at t. We can use the estimated models for the bivariate returns in order to simulate Y_{t+1} and estimate the corresponding VaR and ES. For this purpose we are going to use the two best models from last section with the estimated parameters. They are:

Model 1: univariate margins modelled by a t–distribution and copula mixture of Gumbel and survival Gumbel.

Model 2: univariate margins modelled by a t–distribution and dependence modelled by a t–copula.

The portfolio return distribution has to be approximated by simulation as an exact parametric expression is only known for few particular models. Thus, in order to estimate the portfolio risk measures, first we have to simulate N pairs of returns. In Embrechts et al. [29] we can see how to simulate multivariate observations from a copula–based model. Denote the ith simulated pair by $(x^*_{1,i}, x^*_{2,i})$. Then compute from (1.30) the ith simulated portfolio return for time $t+1$, $y^*_{t+1,i}$. From the N simulated $y^*_{t+1,i}$ we estimate the α Value–at–Risk for the next period, VaR_{t+1} using the empirical quantile given by

$$\mathrm{VaR}_{t+1} = -y^*_{t+1,(\alpha(N+1))}, \tag{1.31}$$

where $y^*_{t+1,(r)}$ is the rth sorted value in ascending order. The expected shortfall estimate for moment $t+1$ will be computed as

$$\mathrm{ES}_{t+1} = -\frac{1}{N_{t+1}} \sum_{i=1}^{N} y^*_{t+1,i}\, \mathbb{I}_{\{x\in\mathbb{R}\,:\,x\leq -\mathrm{VaR}_{t+1}\}}(y^*_{t+1,i}), \tag{1.32}$$

where N_{t+1} is the number of replicated returns $y^*_{t+1,i}$ with value less or equal to $-\mathrm{VaR}_{t+1}$, i.e. $N_{t+1} = \#\{y^*_{t+1,i} : y^*_{t+1,i} \leq -\mathrm{VaR}_{t+1}, i = 1, 2, \ldots, N\}$. We must say here that, from the N simulated $y^*_{t+1,i}$ portfolio returns we can also estimate the VaR_{t+1}, the ES_{t+1} and corresponding confidence intervals using the Peaks Over Threshold (POT) method from Extreme Value Theory instead of (1.31) and (1.32). This

Model	Level	VaR_{t+1} ($\hat{s.e.}$)	ES_{t+1} ($\hat{s.e.}$)
Model 1	90%	0.0177 (0.0011)	0.0272 (0.0017)
	95%	0.0239 (0.0015)	0.0339 (0.0026)
	99%	0.0393 (0.0044)	0.0512 (0.0074)
Model 2	90%	0.0176 (0.0011)	0.0274 (0.0017)
	95%	0.0243 (0.0016)	0.0342 (0.0026)
	99%	0.0395 (0.0046)	0.0518 (0.0076)

Table 1.4: *One week VaR and ES estimates for moment $t+1$ obtained with Models 1 and 2. The bootstrap estimates for the standard errors are also provided.*

method is detailed in Embrechts et al. [27] and the estimates can be obtained using the EVIS software from A. J. McNeil, available on–line at `http://www.math.ethz.ch/~mcneil`. The POT method is potentially better for very high quantiles, like 99.9% or more. Here we follow the non–parametric approach using (1.31) and (1.32). While in the parametric POT method the standard errors come naturally from the maximum likelihood approach, with the empirical quantile method bootstrapping has to be used.

Based on $N = 100\,000$ simulations from Model 1 and the same number from Model 2 we estimated VaR_{t+1} and ES_{t+1} for the confidence levels 90%, 95% and 99%. The results are given in Table 1.4.

The accuracy of the obtained estimates for VaR_{t+1} and ES_{t+1} depends on the accuracy of the model parameter estimates. In simulating the N pairs $(x^*_{1,i}, x^*_{2,i})$ of returns from Model 1 and 2 we use the estimated parameters listed in Tables 1.1 and 1.3. Hence, we obtain point estimates for VaR and ES which accuracy depends on the variability coming from the estimation of the model parameters. To infer about the variance of the model parameter estimates we computed n leave–one–out jackknife estimates of each model parameter and used Proposition 1.17. In order to take the model estimation error into account in the empirical VaR estimate (1.31) we can not use the jackknife method again. It is known that the jackknife estimator for the variance of quantiles is inconsistent, see [22, Chapter 3] and [76, Chapter 2]. An alternative is to compute the bootstrap variance es-

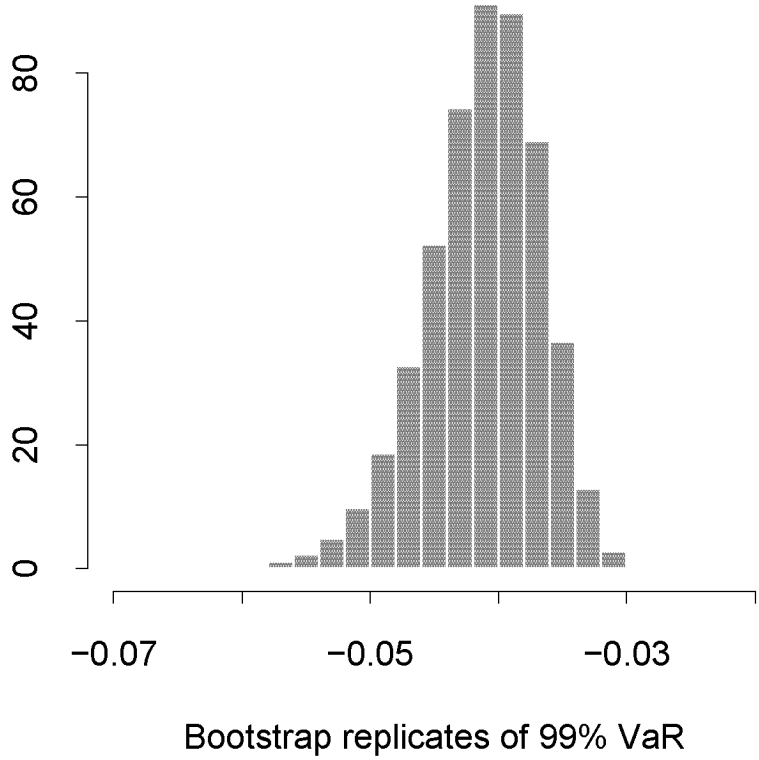

Figure 1.7: *Bootstrap replicates of the one week* 99% *VaR with Model 1 for the portfolio of* 50% *of USD/DEM and* 50% *of USD/JPY.*

timator with Monte Carlo simulation. The method consists of simulating from the estimated model B samples of size n and compute the bootstrap replicate $\text{VaR}^*_{t+1,b}$ for each simulated sample $(\mathbf{x}^*_{1,b}, \mathbf{x}^*_{2,b}, \ldots, \mathbf{x}^*_{n,b})$ for $b = 1, 2, \ldots, B$. With the bootstrap VaR copies we approximate the standard error through

$$\hat{se}(\text{VaR}_{t+1}) = \left(\frac{1}{B} \sum_{b=1}^{B} \left(\text{VaR}^*_{t+1,b} - \frac{1}{B} \sum_{i=1}^{B} \text{VaR}^*_{t+1,b} \right)^2 \right)^{1/2}. \quad (1.33)$$

This is more computer intensive than the jackknife method but besides variance estimation, the bootstrap VaR replicates $\text{VaR}^*_{t+1,b}$ can be used to provide accurate confidence intervals as they give the distribution of VaR_{t+1}. Figure 1.7 plots the histogram of the 99% VaR copies obtained with Model 1.

We used the same method to estimate the standard errors for ES_{t+1}. A number of $B = 100\,000$ samples of size $n = 649$ were simulated from

each model in order to produce the approximate standard errors given in Table 1.4.

Both models give very similar results for VaR and ES. If we obtained significantly different results, we should rethink our models. Not surprisingly, the errors increase as the estimated risk measures go further in the tail.

This approach relies strongly on computational power and estimates are obtained by running a considerable number of simulations; in doing so, we assumed that the error due to the simulation procedure is negligible compared with the error coming from the estimation of the model parameters. Hence, we ignore the error coming from approximating through simulation the distribution of the portfolio value given by the model. For example, for the 99% VaR from Model 1 we get a standard error for the simulation procedure of 0.0003 from estimating this quantile running $N = 100\,000$ simulations. This gives an idea about the magnitude of the simulation error ignored here. It is important to note however that this type of error can be reduced by increasing the number of simulations while the standard errors given in Table 1.4 always remain of the same order.

1.7.3 Backtesting VaR and ES

If we want to backtest the models we can use part of the information available to estimate the one period ahead risk measures and afterwards check them with the values actually observed. For example we can take the weekly returns up to the end of 1990, defining a window of 244 observations, and use these to estimate the weekly VaR and ES to the first week of 1991. Next, we move the data window one week including the observed returns of the first week of 1991 and dropping the oldest bivariate return. Then, we estimate the VaR and ES for the second week of 1991. Proceeding like this, moving the window of data and estimating one week ahead at the time we obtain a series of one week estimations which can be compared with the values actually observed. This procedure, illustrated here for the estimated static models, can readily be used with the more complex time–varying models of Chapters 3 and 4.

Figure 1.8 plots the observed returns of a portfolio of USD/DEM and

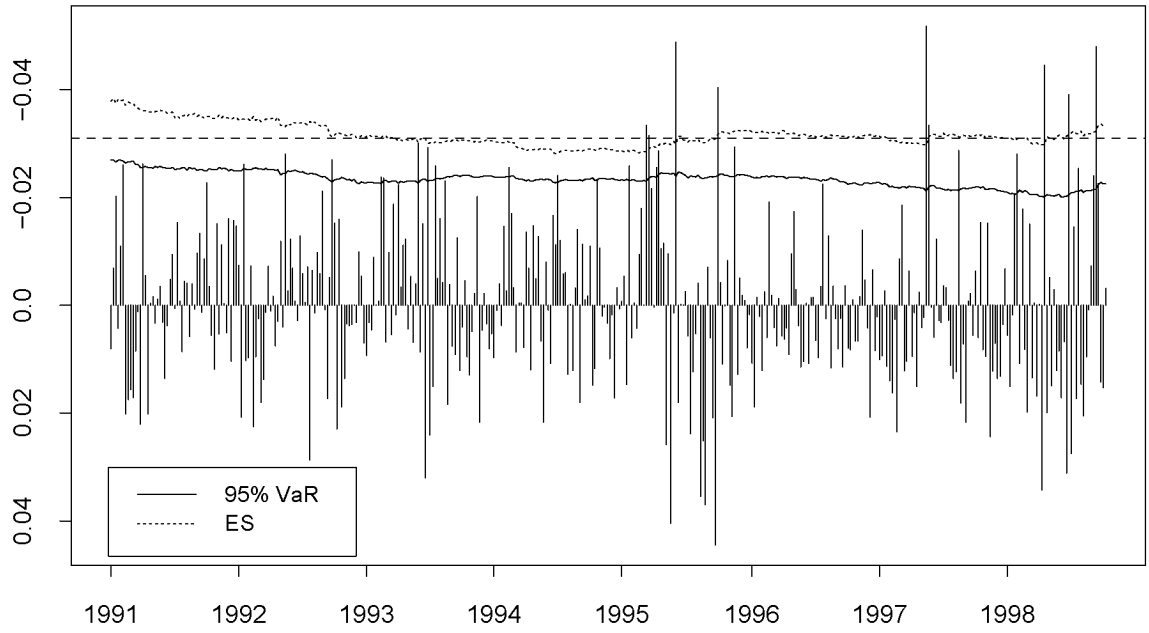

Figure 1.8: *Portfolio weekly returns between January 6, 1991 and October 4, 1998. The portfolio has constant composition of half USD/DEM and half USD/JPY. Model 1 is used to estimate the* 95% *weekly VaR and ES plotted. There are* 7.1% *of violations of estimated VaR values. The long–dashed line is the sample mean of the observed return violations.*

USD/JPY with weights 0.5 each, assuming that this composition remains constant in time. We plotted in the same figure the one week ahead VaR and ES estimates for a level of 95%. Note that the y axis is inverted in the plot. The losses and the gains are switched. VaR and ES are estimated using Model 1 from the previous section. We model the univariate margins by a t–distribution and the dependence with a mixture of Gumbel and survival Gumbel copulae. Assuming iid observations this model ignores time dependence and, more importantly in this case, ignores changing and dependent serial volatility. These questions will be considered later (Chapters 3 and 4); for now this application is suitable to illustrate the IFM method. The static approach to VaR estimation, namely historical simulation, is very well known and widely used in practice, although with less sophistication in the dependence modelling. Actually, analysing the results obtained with Model 1 we obtain that the estimated 95% VaR is violated by 7.1% of the observations. So the point estimates of the model are underestimating the VaR. However, a confidence band surely contains

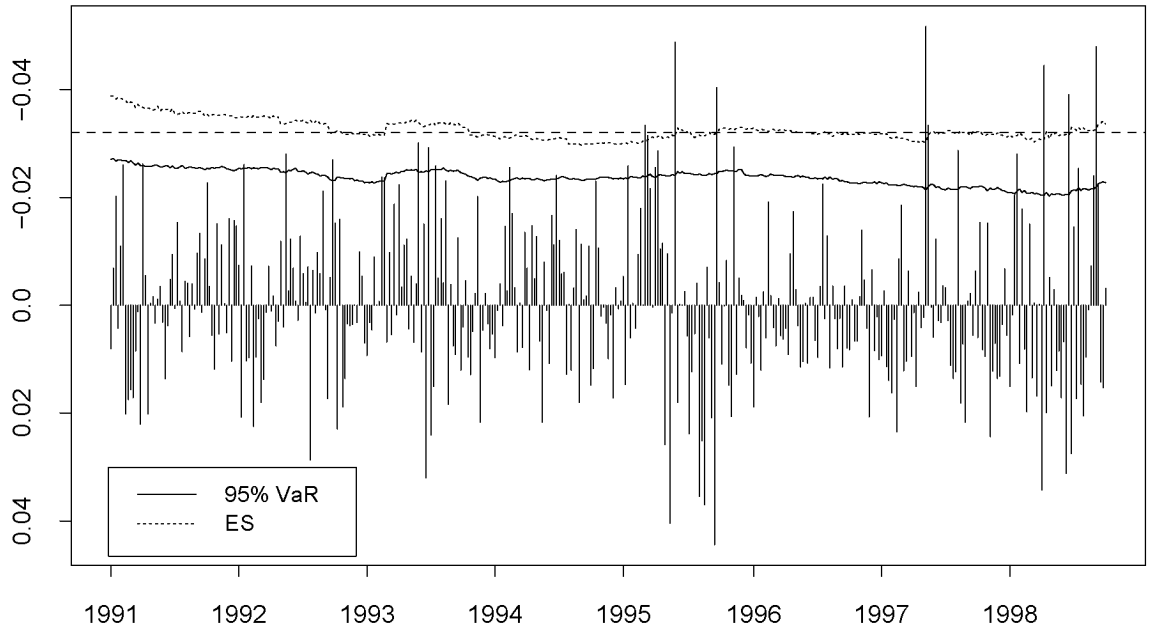

Figure 1.9: *Portfolio returns between January 6, 1991 and October 4, 1998. The portfolio has constant composition of half USD/DEM and half USD/JPY. Model 2 is used to estimate the* 95% *weekly VaR and ES plotted. There are* 6.9% *of violations of estimated VaR values. The long–dashed line is the sample mean of the observed return violations.*

the 5% level of violations. Except in the two first years, the estimated ES is close to the sample mean of the violating observations that surpassed the estimated VaR; see the dotted (estimated ES) and the long–dashed (sample mean of violations) lines in plot of Figure 1.8.

For the same period and portfolio composition we estimated the 95% weekly VaR and ES using Model 2. The margins are now modelled by a t–distribution and the dependence by a t–copula. The results obtained are very much the same as those given by Model 1. There are 6.9% of observed returns surpassing the VaR and the sample mean of these VaR violations is close to the estimated ES. Figure 1.9 plots the results for Model 2. For the 99% VaR case (not plotted here) Model 1 and Model 2 gave respectively 1.7% and 1.5% of returns over the estimated VaR. Hence, the point estimates from the two models underestimate the VaR also for higher quantiles. In practice we would provide all the estimates with bootstrap confidence intervals using (1.33). Even without having done

these calculations, given that the point estimates are so close to the target levels, we conclude that in this case the copula–based models are capturing quite well the risk measures selected.

1.7.4 Tail–dependence coefficient estimation

The fitted models can also be used to estimate the tail–dependence coefficient from Definition 1.12. As this coefficient is a copula feature we only need the copula function from each model in order to estimate it. Moreover, the univariate margins in Models 1 and 2 are continuous (we assumed a t–distribution). Then, by Proposition 1.14, the returns and the copula of the returns have the same tail–dependence coefficient. Model 1 has a copula function of the form

$$C(u_1, u_2; \boldsymbol{\theta}) = \\ \theta_3 C^{Gu}(u_1, u_2; \theta_1) + (1 - \theta_3)(u_1 + u_2 - 1 + C^{Gu}(1 - u_1, 1 - u_2; \theta_2))$$

where C^{Gu} is a Gumbel copula and $(u_1, u_2) \in [0, 1]^2$. Using the definition of a Gumbel copula (1.4) we obtain that

$$C(u_1, u_1; \boldsymbol{\theta}) = \theta_3 u_1^{2^{1/\theta_1}} + (1 - \theta_3)\left(2u_1 - 1 + (1 - u_1)^{2^{1/\theta_2}}\right). \qquad (1.34)$$

Using (1.34) in Proposition 1.13 *i*) and *iii*) we obtain, respectively, the lower and upper tail–dependence coefficients for Model 1. They are

$$\lambda_L = (1 - \theta_3)\left(2 - 2^{1/\theta_2}\right) \quad \text{and} \quad \lambda_U = \theta_3\left(2 - 2^{1/\theta_1}\right). \qquad (1.35)$$

As far as Model 2 is concerned, the tail–dependence coefficient for the t–copula was given in (1.6).

Table 1.5 has the estimated values for the tail–dependence coefficient and respective standard errors. Here we used (1.16) to estimate the variance of the λ estimators.

Until now the results given by the models have been very similar. With the tail–dependence coefficient this is not so obvious. Model 1 gives point estimates higher than Model 2. Notice that we are estimating an asymptotic tail feature using a model fitted with the entire data set. Hence,

	λ_L $(\hat{s.e.})$	λ_U $(\hat{s.e.})$
Model 1	0.1900 (0.0543)	0.2932 (0.0556)
Model 2	0.1226 (0.0873)	0.1226 (0.0873)

Table 1.5: *Lower and upper tail–dependence coefficients for the weekly returns on USD/DEM and USD/JPY spot rates given by the fitted Models 1 and 2.*

	CI	λ_L	λ_U
Model 1	empirical	[0.0509 , 0.2965]	[0.1723 , 0.4013]
	normal	[0.0835 , 0.2965]	[0.1842 , 0.4023]
Model 2	empirical	[0 , 0.2965]	[0 , 0.2965]
	normal	[0 , 0.2938]	[0 , 0.2938]

Table 1.6: *Jackknife confidence intervals for tail–dependence coefficient estimates constructed using empirical and normal approaches for a* 90% *probability level.*

most of the influence in the fitting process comes from observations in the centre of the distribution. Nevertheless, looking at the estimated standard errors in Table 1.5 we suspect that statistically we might not reject the equality of the tail–dependence coefficients given by the two models. In fact, from the jackknife estimates we can compute normal or empirical confidence intervals for λ_U and λ_L (see Section 1.8). These are reported in Table 1.6 for a 90% confidence level.

In spite of the point estimates for the tail–dependence coefficient given by the two models being apparently different, the (90% level) confidence intervals for these quantities overlap considerably.

We can look at the same question by a slightly different way computing the differences between the tail–dependence coefficient estimates given by the two models. Doing that for the lower and upper tails respectively,

$$\lambda_L(\text{Model 1}) - \lambda(\text{Model 2}) \quad \text{and} \quad \lambda_U(\text{Model 1}) - \lambda(\text{Model 2}),$$

the confidence intervals for these estimators are the ones listed in Table 1.7.

Once again we can not rule out the hypothesis of equality between the

tail coefficients of the two models. After this confidence interval analysis we conclude that Model 1 and Model 2 do not give significantly different results about the tail–dependence coefficient, especially in the lower tail. But there is another observation to be made concerning the results in Table 1.6. The confidence intervals for the tail coefficient in Model 2 include the zero value. This means that Model 2 does not exclude the hypothesis of asymptotic independence in the tails. In practice however multivariate high (low) observations seem to occur too often for this hypothesis to be true, hence a model focused on the multivariate tails should be more appropriate to tail dependence inference. Also, it is observed in practice that extremal clustering occurs in periods of crisis so that in order to study such phenomena a more dynamic modelling has to be taken into account. Finally, when one is really interested in specific properties of joint tails, one should use estimation procedures which are based solely on joint upper or lower tails. We will come back to these issues in Chapters 2 and 3.

CI	λ_L(Mod.1) − λ(Mod.2)	λ_U(Mod.1) − λ(Mod.2)
Empirical	[-0.1305 , 0.2358]	[-0.0768 , 0.3499]
normal	[-0.1287 , 0.2635]	[-0.0287 , 0.3699]

Table 1.7: *Jackknife* 90% *confidence intervals for the difference between tail–dependence coefficient estimates obtained by Model 1 and Model 2.*

We finally would like to remark that tail–dependence estimation as explained above is widely used in empirical finance in order to study systemic risk and spillover effects; see for instance Borio et al. [6] and Hartmann et al. [44].

1.8 A note on jackknife confidence intervals

As we have the jackknife estimates for the model parameters and consequently also have the jackknife estimates for the tail–dependence coefficients it is straightforward to use these in constructing confidence intervals. The normal confidence intervals are built on the assumption that $\hat{\lambda}$ is a centred, approximately normally distributed estimator with standard

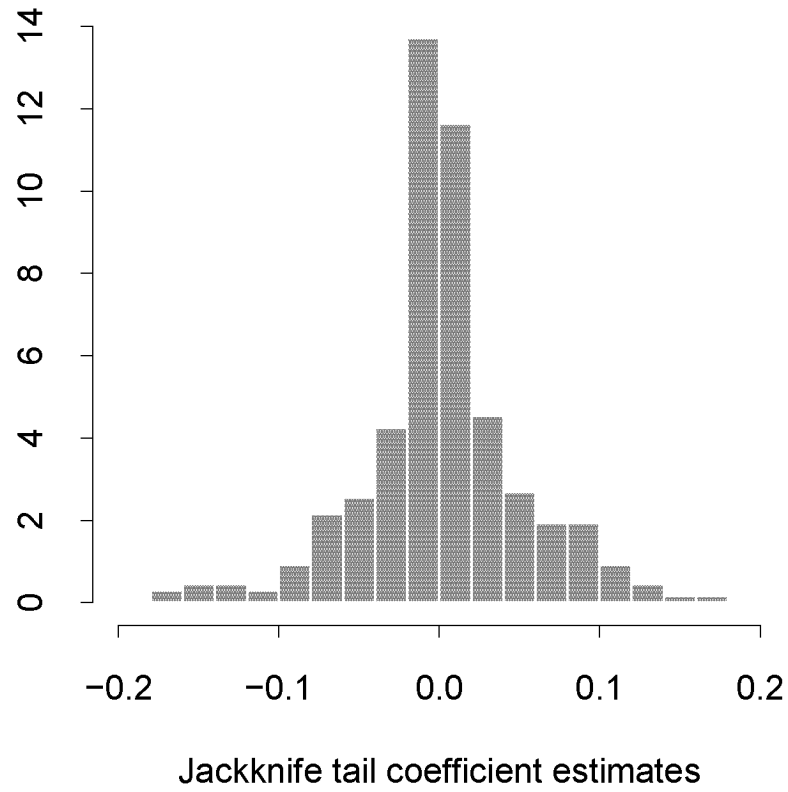

Figure 1.10: *Inflated jackknife values, centred at the jackknife mean, of the lower tail–dependence coefficient of weekly returns USD/DEM and USD/JPY estimated with Model 1.*

deviation given by the jackknife error estimate $\hat{s.e.}(\hat{\lambda})$. Thus the $1-\alpha$ confidence interval for $\hat{\lambda}$ is given by

$$\left[\hat{\lambda} - \hat{s.e.}(\hat{\lambda})q_{1-\alpha/2}, \hat{\lambda} + \hat{s.e.}(\hat{\lambda})q_{1-\alpha/2}\right],$$

where q_α is the α–quantile of a univariate standard normal distribution. In case $\hat{\lambda}$ does clearly not have an approximate normal distribution these confidence intervals are not very accurate. To overcome that possibility we can construct empirical confidence intervals using the so–called inflated jackknife values; see [23, Chapter 11]. Suppose that from a sample $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ of size n we estimate the leave–one–out jackknife quantities $\hat{\lambda}^{(-i)}$ for $i = 1, 2, \ldots, n$. The inflated jackknife values, centred at the jackknife mean, are given by

$$\sqrt{n-1}\,(\hat{\lambda}^{(-i)} - \hat{\lambda}^{(\cdot)})$$

where

$$\hat{\lambda}^{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\lambda}^{(-i)}.$$

Figure 1.10 displays a histogram of these values for the lower tail–dependence coefficient of Model 1. The two sided $1-\alpha$ empirical confidence interval for the estimator $\hat{\lambda}$ used is given by

$$\left[\hat{\lambda}+\sqrt{n-1}\left(\hat{\lambda}^{(-i)}_{(\alpha/2(n+1))}-\hat{\lambda}^{(\cdot)}\right),\hat{\lambda}+\sqrt{n-1}\left(\hat{\lambda}^{(-i)}_{((1-\alpha/2)(n+1))}-\hat{\lambda}^{(\cdot)}\right)\right]$$

where $\hat{\lambda}^{(-i)}_{(r)}$ is the rth value from the set $\left\{\hat{\lambda}^{(-i)} : i=1,2,\ldots,n\right\}$ sorted by ascending order.

Comments

In this chapter we have used maximum likelihood to estimate the copula parameters. This is applicable to any copula family. In some special cases other approaches are possible. Whenever there exists a known one-to–one mapping between the copula parameter θ and Kendall's τ (see [35]), θ can be estimated as a function of the usual Kendall τ estimate. There, the properties of the copula parameter estimator depend on the properties of the Kendall τ estimator. Throughout the chapter we used the AIC as a model selection criterion. Genest and Rivest [38] give an alternative method for a particular class of copulae referred to as Archimedean copulae (see [63, Chapter 4]).

Univariate data in finance and insurance often exhibit heavy–tail behaviour. The methods we introduced focused more on the dependence structure rather than on the modelling of univariate margins. If a very accurate modelling of the univariate tails plays an important role in the final analysis, a univariate semi–parametric approach may improve the modelling of data with heavy tails. We can replace the first step of the IFM or pseudo log–likelihood methods with this alternative. There, the bulk of the data is modelled by the empirical distribution function, say, and the tail by a generalised Pareto distribution (GPD) fitted with maximum likelihood. Details on inference methods for GPD are given in Embrechts et al. [27, Chapter 6].

With the tools presented, we were able to perform a statistical analysis of multivariate data in finance. In particular we used copula fitting and applied the models selected to particular problems in financial risk management: VaR and ES estimation and backtesting. In the following chapters we shall concentrate on the dynamic modelling of the dependence structure of financial (FX) data. We refrain from working out similar risk management applications, though these could easily have been provided.

Chapter 2

Stationary copula analysis

Stylised econometric facts of one–dimensional financial (FX) returns are well studied. The work by Olsen & Associates has extended these facts across sampling frequencies reaching from minutes to months. The most detailed summary for tick–by–tick (high–frequency) FX data is to be found in [15]. Much less results are known for multivariate series. The main reason for this is the lack of modelling tools specifically developed with this goal in mind. The copula methods introduced in Chapter 1 may provide such a tool. Clearly, any serious analysis must take time variation into account; we will do so progressively in the next chapters. Our basic data set is the bivariate time series of FX spot rates (USD/DEM, USD/JPY) already used in Chapter 1 at the level of weekly returns. In this chapter we will combine detailed econometric modelling of the high–frequency data set at hand with specific applications of copula based methods which pay special attention to the behaviour of bivariate extremes. Our final goal is to obtain a better understanding of the dependence structure of the FX returns across different time frequencies. In several instances, we will explicitly use the richness of the high–frequency structure. This chapter can also be seen as a road map for future research on the analysis of multivariate tick–by–tick data in finance.

The outline of the chapter is as follows. Section 2.1 presents a method for deseasonalising the returns for time horizons up to one day. The

raw high–frequency (tick–by–tick) data is then transformed to properly deseasonalised data. In Section 2.2, several families of copulae will be fitted to deseasonalised two–dimensional FX data, and this at several frequencies (from hourly to daily). Goodness–of–fit tests, including tests for ellipticity, are performed in Section 2.3. In Section 2.4, the problem of clustering of bivariate extremes in the data is analysed.

2.1 Deseasonalisation of the returns

After being collected by Olsen Data, the observations (tick–by–tick FX quotes) are filtered removing transmission errors, fake quotes originated by transmission tests, etc. A description of this filtering process is to be found in Dacorogna et al. [15]. The FX data set available covers the period from April 1986 until October 1998.

Originally the number of quotes emitted by the FX market is very high (around ten million for USD/DEM in the given period) and they are irregularly spaced in time. Regular time series are obtained reducing the observations to a time step δ of five minutes. Missing observations are obtained by linear interpolation. This is one of the standard techniques to approach the missing values question in time series analysis; see Schafer [74] and references therein. For a given currency, a single quote at time t consists of a bid price, $p_{t,\text{bid}}$ and an ask price, $p_{t,\text{ask}}$. As we are not interested in the effects related to the bid–ask spread, we consider logarithmic middle prices. The logarithmic middle price for time t is defined as

$$\bar{p}_t = \frac{1}{2}\left(\log p_{t,\text{bid}} + \log p_{t,\text{ask}}\right).$$

These are displayed in Figure 2.1 for the USD/DEM and USD/JPY spot rates. Naturally, the logarithmic return at t with respect to the time horizon (or frequency) Δt, $\tilde{x}_{t,\Delta t}$ is given as the difference of logarithmic middle prices

$$\tilde{x}_{t,\Delta t} = \bar{p}_t - \bar{p}_{t-\Delta t}. \tag{2.1}$$

One hour logarithmic returns are displayed in Figure 2.2. The definition in (2.1) is for a general time horizon Δt. Nevertheless, in this chapter, whenever we refer to five minutes returns we drop the frequency subscript Δt.

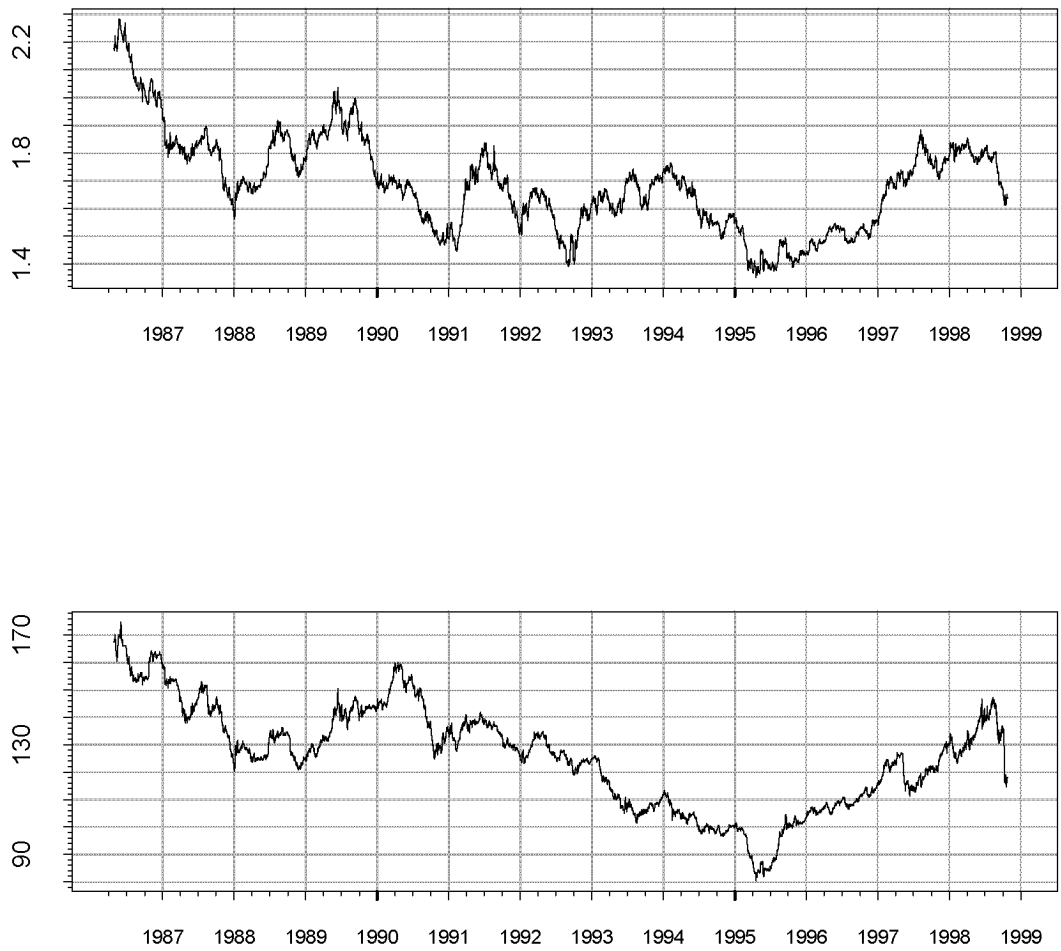

Figure 2.1: *Logarithmic middle prices for USD/DEM (top) and USD/JPY (bottom) spot rates.*

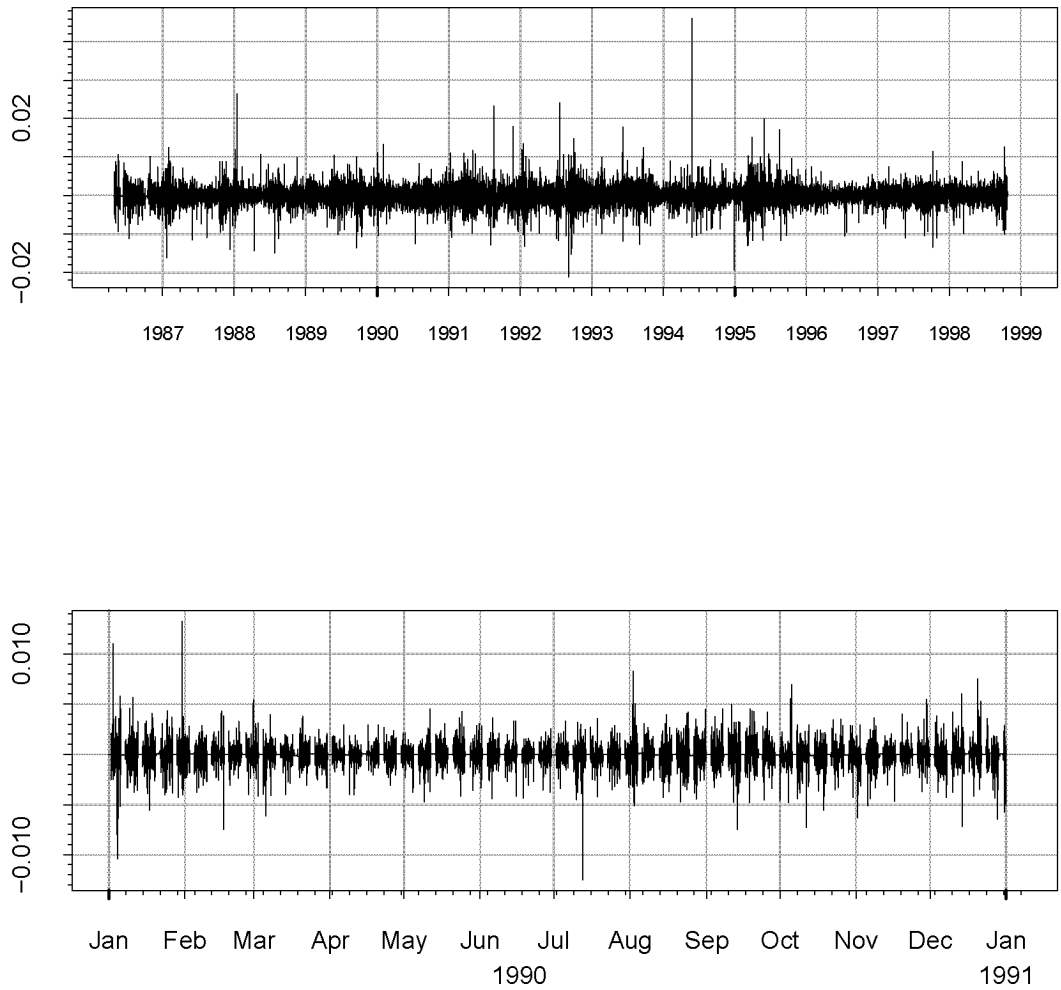

Figure 2.2: *Hourly returns on USD/DEM for the whole sample in the top panel and for one year in the bottom one. Weekly seasonality is visible in the bottom panel.*

The low activity in the FX market during the weekends is visible in the returns plotted in the bottom panel of Figure 2.2. Hence, the changing market activity induces a cyclic behaviour on the variability of the returns. In other words, the market volatility possesses a seasonal component, the so–called seasonal volatility pattern. This is very pronounced in high–frequency data. In Figure 2.3 we can clearly distinguish daily and weekly cycles in the sample autocorrelation and cross–correlation functions of absolute values of the hourly returns. The classical theory of time series modelling views seasonalities as some kind of data contamination that should be eliminated through a deseasonalisation method; see Brockwell and Davis [9].

Specific approaches to the deseasonalisation of high–frequency data can be found in the literature; at a text book level see Dacorogna et al. [15]. These methods basically fall into two categories: time transformation and volatility weighting. Time transformation methods consist of changing from the physical to an activity–related time scale. The modelling of the activity time scale of an FX rate depends on its volatility. Consequently, a different activity time scale is found for different FX rates. Component–wise time transformation of simultaneous multivariate observations leads eventually to desynchronised observations. This fact rules out a univariate time scale transformation approach in multivariate time series.

The alternative deseasonalisation method, referred to as volatility weighting, consists of standardising the time series of the returns by a volatility estimated conditional on the time of the week. Formally, denote by $\widetilde{X}_t$ the log–return at time t for the time frequency of five minutes. Suppose that the return time series data is a realisation of the process

$$\widetilde{X}_t = \mu + v_t \, X_t, \tag{2.2}$$

where μ is the mean log–return, v_t is the five minutes expected volatility at t and X_t is a random component corresponding to a deseasonalised log–return at moment t for the five minutes frequency. We assume in (2.2) that the trend component of the process is constant. In fact, this is a common feature in financial log–returns and we can see that for the USD/DEM case in Figure 2.2. The expected volatility depends on the market activity as we can observe in Figure 2.2 for the weekends. Hence, the market activity at moment t, a_t has to be taken into account when computing

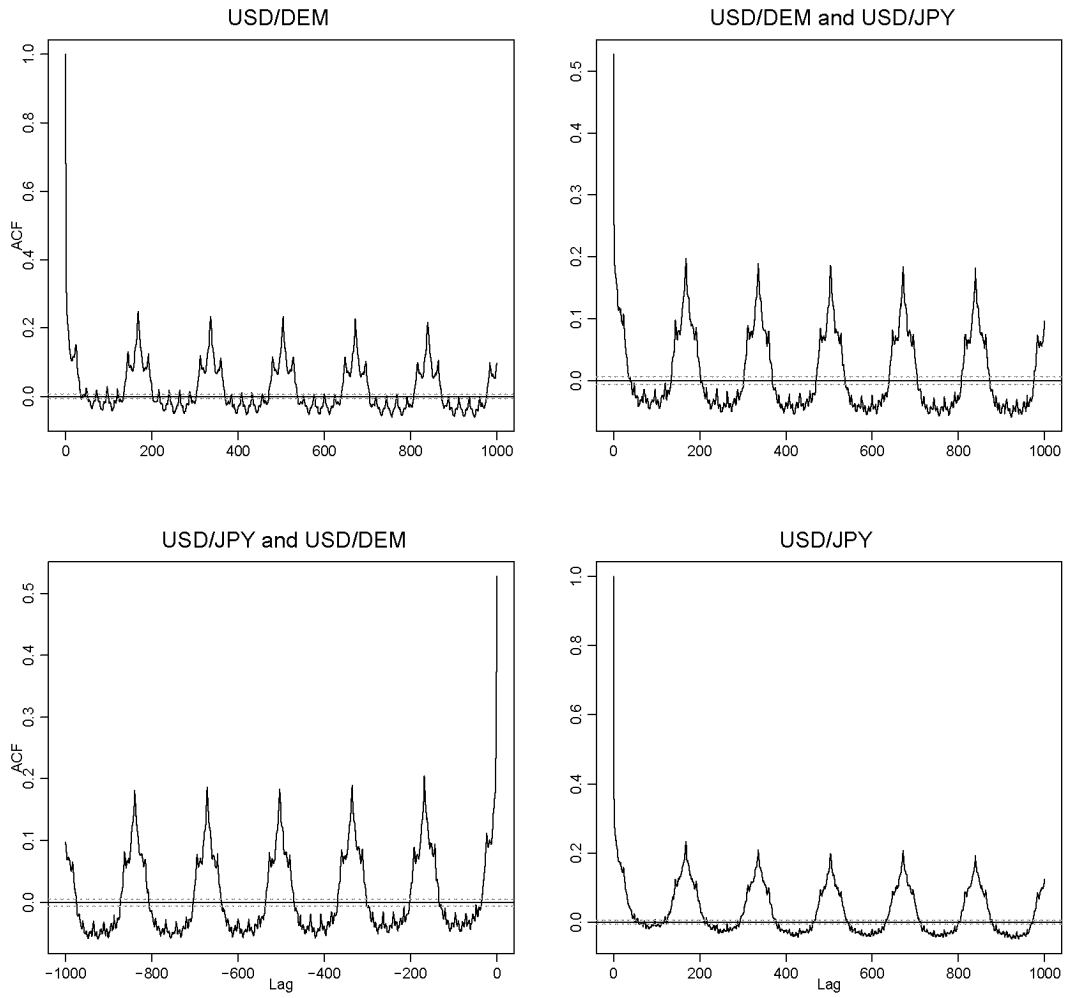

Figure 2.3: *Sample autocorrelation and cross–correlation functions of the absolute values of the hourly returns on USD/DEM and USD/JPY.*

the expected volatility. Working with USD/DEM and USD/JPY we have to consider the activity of the European, the American and East Asian markets and split a_t in three components, one for each market:

$$a_t = a_{t,\text{America}} + a_{t,\text{East Asia}} + a_{t,\text{Europe}}.$$

In a normal day of market activity a_t is one. If a given t_0 falls on a public holiday, in Europe say, then $a_{t_0,\text{Europe}} = 0$ and a_{t_0} is less than one. Moreover, the expected volatility is subject to the shift of daylight saving time (DST) periods. Hence, we have to compute separately the expected volatility for winter and for summer DST periods. We have to estimate the volatility conditional on the moment of the week because there is a weekly and intra–day seasonality; see Figure 2.3. A moment t can be written in the form

$$t = t_i + \tau\delta$$

where t_i is the beginning of week i (Sunday 00:00:00 GMT) and $\tau \in \{0, 1, \ldots, T\}$ with T the number of five minutes periods δ in a seven days week. Taking these facts into consideration, v_t is modelled as

$$v_t^2 = a_t \cdot \left(v_{\tau\delta}^{(d)}\right)^2, \tag{2.3}$$

where $v_{\tau\delta}^{(d)}$ is the expected volatility at moment $\tau\delta$ of the week in the DST period d, winter or summer. Then, $v_{\tau\delta}^{(d)}$ is estimated as

$$\left(\bar{v}_{\tau\delta}^{(d)}\right)^2 = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\tilde{x}_{t_i+\tau\delta}\right)^2,$$

where N_d is the number of weeks in the DST period d. In Figure 2.4 we plot the volatility patterns $\bar{v}_t$ (estimates of v_t in (2.3)) for the five minutes USD/DEM and USD/JPY time series.

Using the volatility patterns estimated by (2.3) we can compute the five minutes deseasonalised log–returns

$$x_t = \frac{\tilde{x}_t - \bar{x}}{\bar{v}_t}, \tag{2.4}$$

where $\bar{x}$ is the sample mean of the log–returns $\tilde{x}_t$ which were computed from (2.1) with $\Delta t = \delta$. For simplicity and without loss of generality

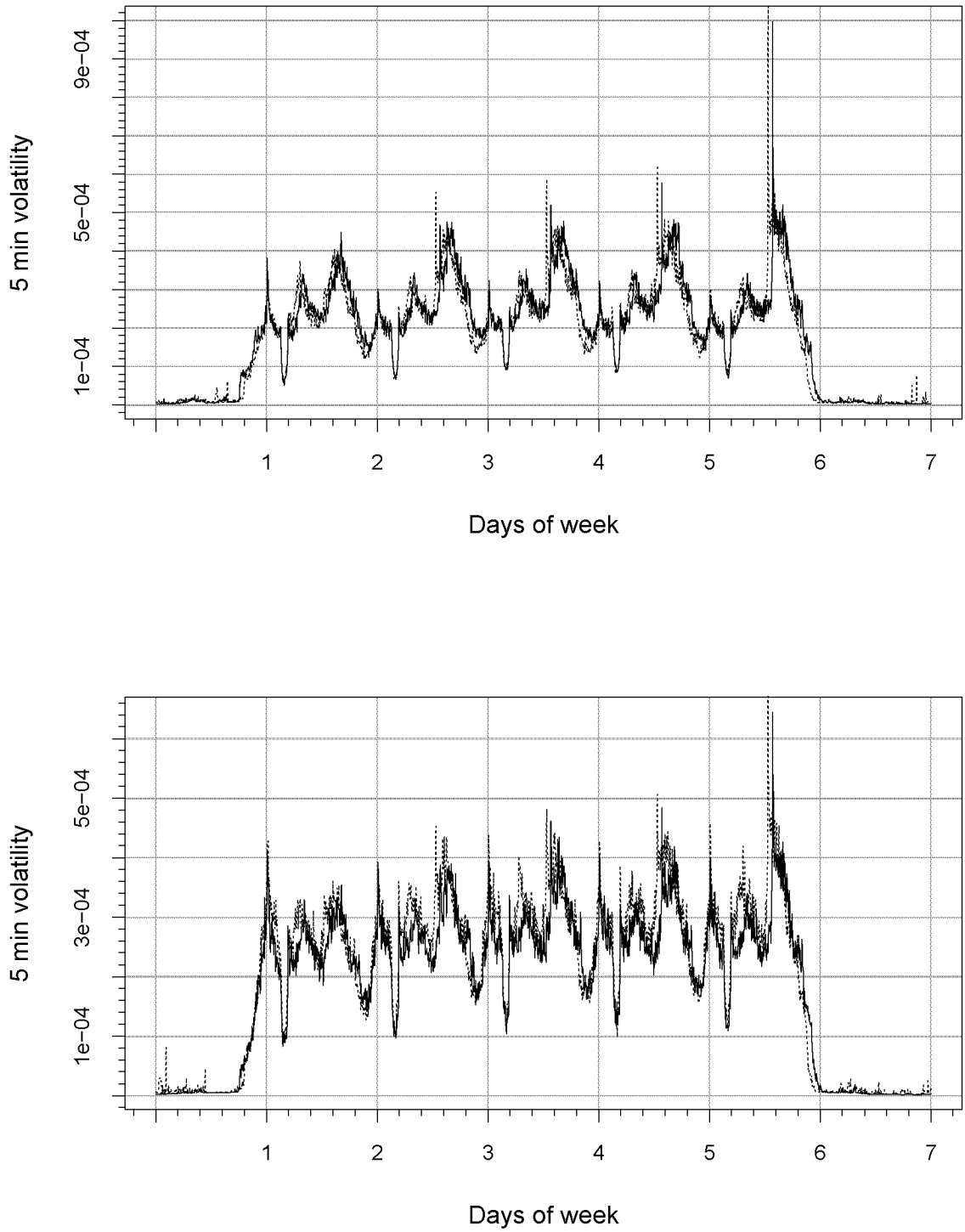

Figure 2.4: *Weekly volatility patterns computed from the USD/DEM (top) and USD/JPY (bottom) time series. Full lines give the winter mean volatility and dotted lines are the summer volatility patterns.*

assume in the following that the returns $\widetilde{X}$ are already centred by their mean μ. Now we need to know how to obtain the deseasonalised returns for any time frequency Δt rather than for five minutes. Consider the simple generalisation of (2.2) from a five minutes to a Δt time frequency process

$$\widetilde{X}_{t,\Delta t} = v_{t,\Delta t}\, X_{t,\Delta t}. \tag{2.5}$$

Note that the log–return process is supposed to be mean centred. The expected volatility at time t for an arbitrary time frequency Δt is defined in terms of the five minutes volatility as

$$v_{t,\Delta t} = \left(\sum_{i=0}^{n-1} v_{t-i\delta}^2 \right)^{1/2},$$

where $n\delta = \Delta t$. Denote by $\overline{P}_t$ the logarithmic middle price at time t. Then, (2.5) can be rewritten in the form

$$\overline{P}_t - \overline{P}_{t-\Delta t} = \left(\sum_{i=0}^{n-1} v_{t-i\delta}^2 \right)^{1/2} X_{t,\Delta t}, \tag{2.6}$$

with $n\delta = \Delta t$. Using (2.6) we can compute the deseasonalised log–returns for any time frequency Δt through

$$x_{t,\Delta t} = \frac{\bar{p}_t - \bar{p}_{t-\Delta t}}{\left(\bar{v}_t^2 + \bar{v}_{t-\delta}^2 + \ldots + \bar{v}_{t-(n-1)\delta}^2 \right)^{1/2}}, \tag{2.7}$$

as a function of the logarithmic middle prices and the five minutes volatility pattern estimated from the data.

The mean activity of the FX market during weekends is very low; recall the plotted returns in the bottom panel of Figure 2.2. The usual approach to handle this cyclic behaviour consists of dropping the weekends from the middle prices series. However, big jumps may appear between closing Friday and opening Monday prices, so that dropping the weekends without a special treatment could produce false big return values. This would then induce spurious seasonality. Therefore, a weekend volatility has to be calculated in order to avoid this. Define the beginning and the end of a weekend respectively as t_{w_0} = Friday, 21:00:00 GMT and t_{w_1} = Sunday,

21:00:00 GMT. Let $\Delta t_w = t_{w_0} - t_{w_1}$ be the weekend length. The expected volatility during weekends is estimated as

$$\bar{v}_t = \left|\tilde{x}_{t_{w_1},\Delta t_w}\right| \left(\frac{\delta}{\Delta t_w}\right)^{1/2}, \tag{2.8}$$

if $t \in (t_{w_0}, t_{w_1}]$. After the volatility pattern estimation is done, the weekends can be dropped in the middle prices series and in the volatility. Because of (2.8), the first volatility estimate of the week is in tune with an eventual price jump during the weekend and no spurious seasonality remains in the deseasonalised returns.

With the volatility patterns plotted in Figure 2.4 we computed the series of deseasonalised log–returns on USD/DEM and USD/JPY for six time frequencies: one, two, four, eight, twelve hours and one day. The one hour deseasonalised log–returns are plotted in Figure 2.5. The sample autocorrelation and cross–correlation functions for the absolute values of two deseasonalised hourly return series are displayed in Figure 2.6 and are to be compared with those in Figure 2.3. The removal of the seasonalities in the two series is obvious.

Furthermore, we plotted in Figures 2.7 and 2.8 the QQ–plots of the normal against the empirical quantiles of the deseasonalised returns on USD/DEM and USD/JPY for the six frequencies considered. As it is usually found in the econometric literature, from those two figures we can see that the univariate distributions are clearly heavy tailed for short time horizons and become more thin tailed as the frequency decreases.

2.2 Dependence structure modelling

Given the bivariate deseasonalised returns from (2.7), we now want to analyse the dependence structure at each of the different frequencies: one, two, four, eight, twelve hours and one day. In each case, we will fit parametric families of copulae using the pseudo log–likelihood method from Section 1.4. Recall that this procedure consists, in a first step, of transforming the marginal observations into uniformly distributed vectors using the rescaled empirical distribution functions. Then, the copula pa-

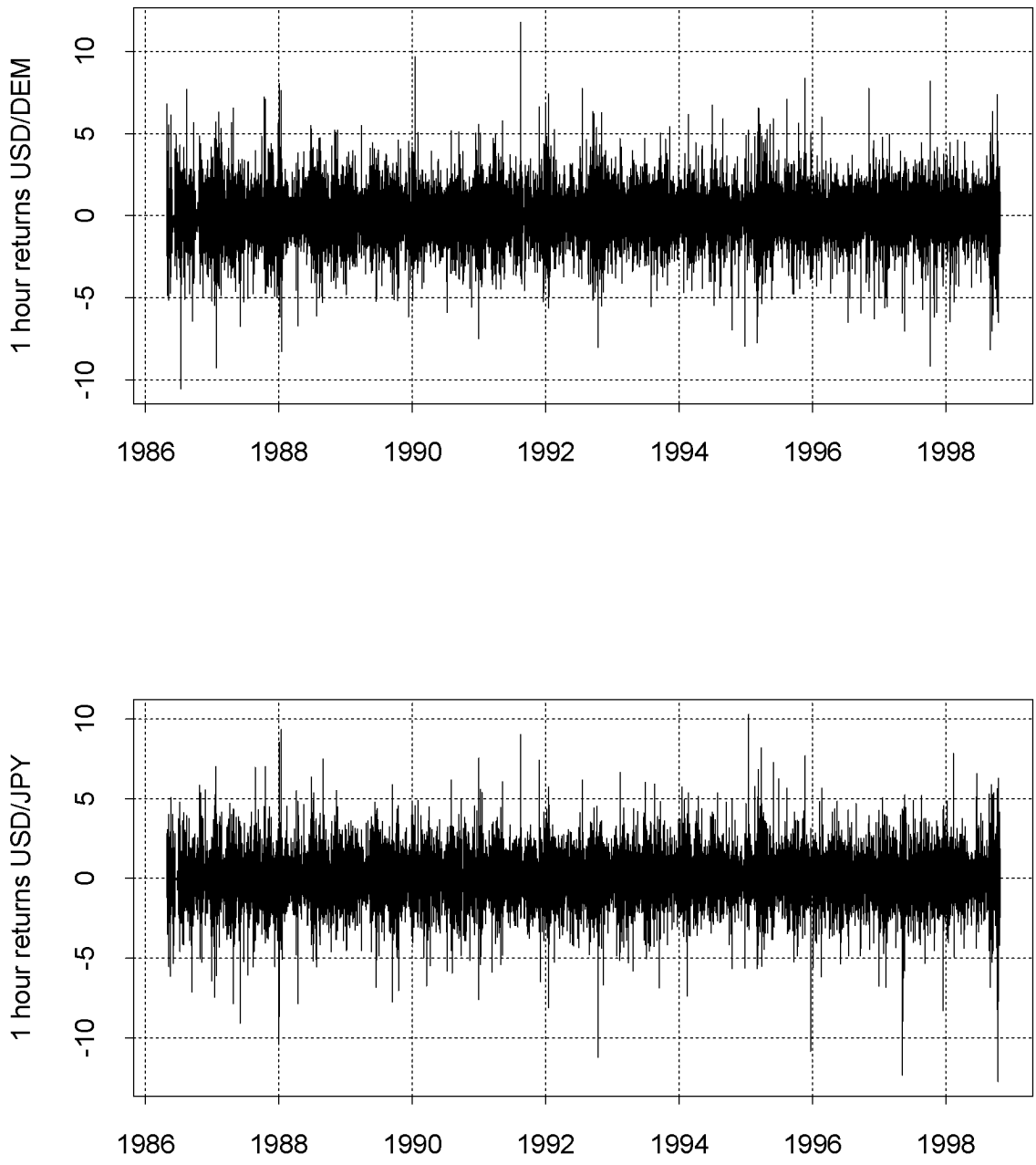

Figure 2.5: *One hour deseasonalised returns on the spot rates USD/DEM and USD/JPY.*

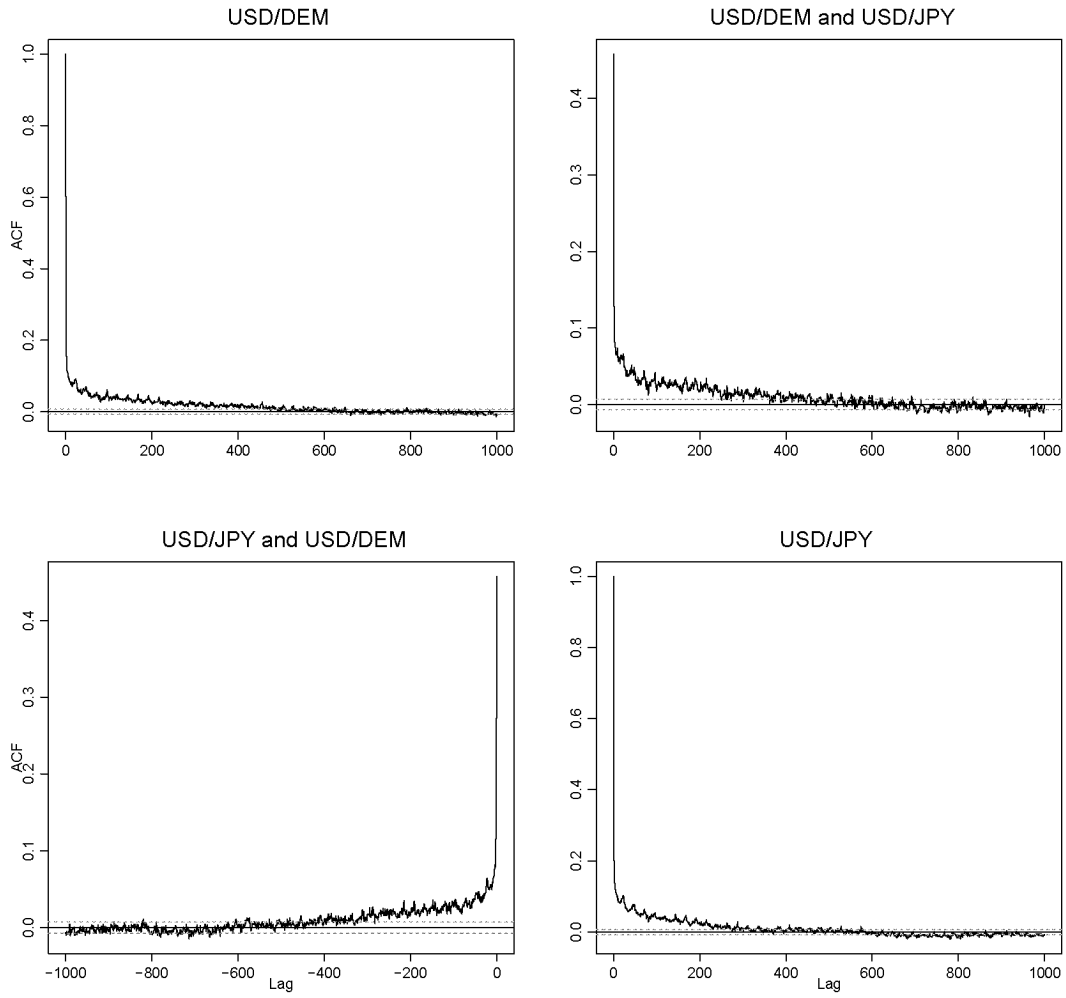

Figure 2.6: *Sample autocorrelation and cross–correlation functions of the absolute values of the deseasonalised USD/DEM and USD/JPY one hour returns.*

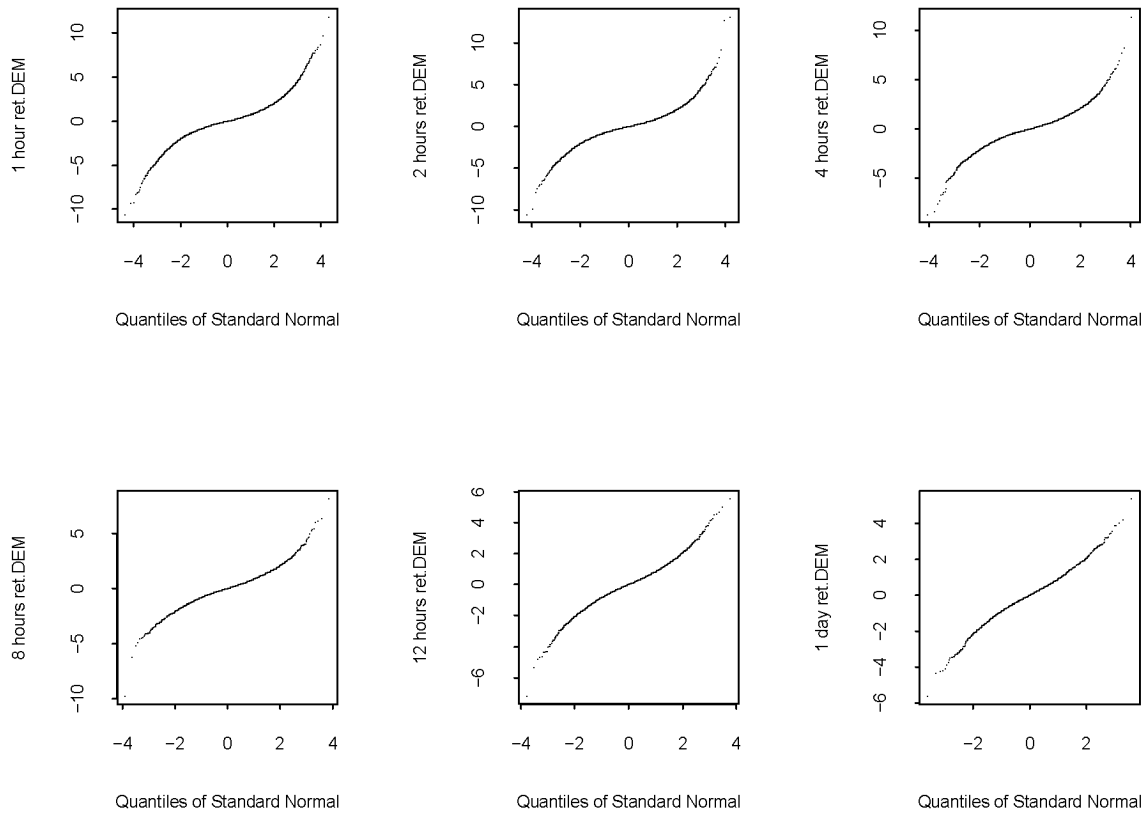

Figure 2.7: *QQ–plots of the normal versus the empirical quantiles of deseasonalised log–returns on USD/DEM spot rate for the six frequencies considered.*

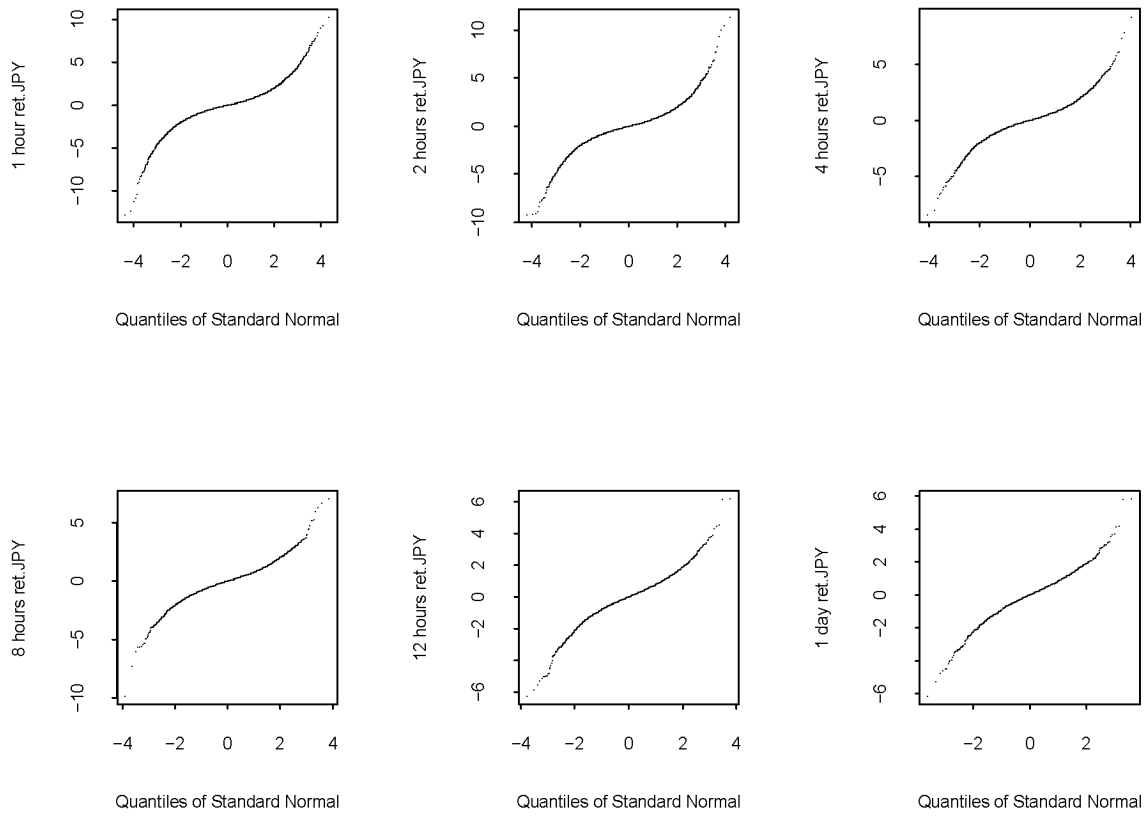

Figure 2.8: *QQ–plots of the normal versus the empirical quantiles of deseasonalised log–returns on USD/JPY spot rate for the six frequencies considered.*

rameters are estimated by the maximisation of the pseudo log–likelihood function (1.18).

For each considered frequency we have two vectors of deseasonalised returns, one for USD/DEM and another for USD/JPY. The corresponding scatter plots are shown in Figure 2.9. We denote the random variable for the deseasonalised USD/DEM returns by X_1 and for the deseasonalised USD/JPY returns by X_2. If $\{x_{i1}, x_{i2}, \ldots, x_{in}\}$ are the n observed univariate deseasonalised returns of the FX rate X_i $(i = 1, 2)$ for a given frequency, then applying

$$F_{in}(x) = \frac{1}{n+1} \sum_{j=1}^{n} \mathbb{I}_{\{y \in \mathbb{R} \,:\, y \leq x\}}(X_{ij}),$$

to the deseasonalised returns we obtain pseudo–returns approximately uniformly distributed in $[0, 1]$. Figure 2.10 displays the scatter plots of the bivariate pseudo–returns

$$(F_{1n}(x_{1i}), F_{2n}(x_{2i})), \tag{2.9}$$

for $i = 1, 2, \ldots, n$ and Figure 2.11 shows the contour–plots of the same returns. To better visualise the several bivariate shapes we can plot the data with univariate standard normal margins at all the six frequencies. Figure 2.12 shows the plots of the transformed pseudo–returns

$$\left(\Phi^{-1}(F_{1n}(x_{1i})), \Phi^{-1}(F_{2n}(x_{2i}))\right)$$

for $i = 1, 2, \ldots, n$, where Φ is the distribution function of a univariate standard normal random variable. The number of points plotted in each panel varies a lot and that makes them harder to compare. But even so, we can still see in Figure 2.12 that there is an evolution from a diamond to an elliptic shape as the time frequency decreases. In Figure 2.10, for the one, two and four hour returns, we plotted sub–samples of the pseudo–returns otherwise the scatter plots of the full samples would be just three useless black squares.

On the transformed data sets (2.9), for each frequency we estimate the parameters for several copula families. Here we consider the Gaussian, the t, the Frank, the Gumbel and the Clayton copulae. The definitions

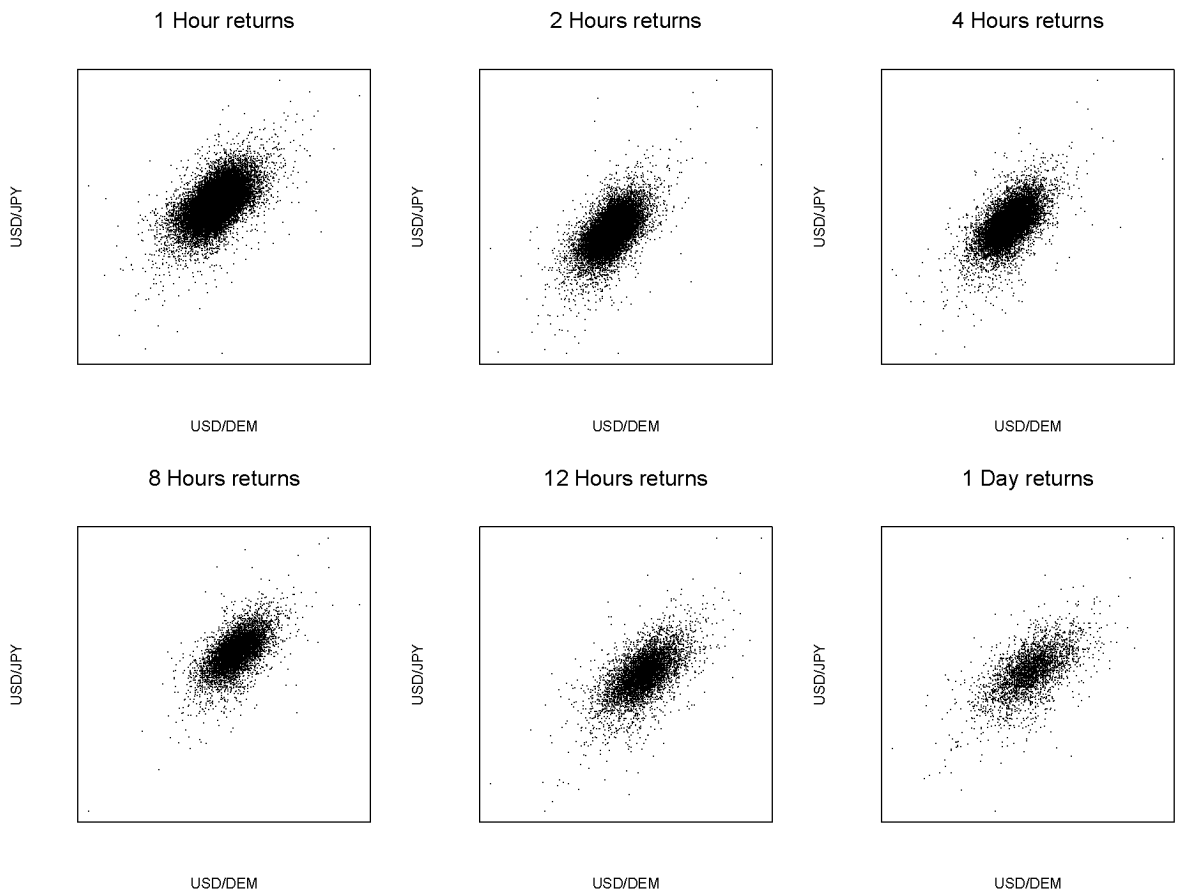

Figure 2.9: *Scatter plots of deseasonalised returns of USD/DEM and USD/JPY for different time frequencies.*

Frequency	n. obs.	Model	Estimate	$\hat{s.e.}$	AIC
1 hour	77 758	t	0.563; 4.339	0.003; 0.087	-32698
		Gumbel	1.577	0.005	-29052
		Gaussian	0.555	0.002	-28674
		Frank	4.030	0.025	-27275
		Clayton	0.880	0.007	-23997
2 hours	38 976	t	0.585; 4.269	0.004; 0.120	-17951
		Gumbel	1.622	0.007	-16066
		Gaussian	0.578	0.003	-15859
		Frank	4.252	0.036	-14983
		Clayton	0.944	0.010	-13267
4 hours	19 514	t	0.599; 4.282	0.005; 0.169	-9481
		Gumbel	1.652	0.009	-8530
		Gaussian	0.592	0.004	-8400
		Frank	4.402	0.051	-7957
		Clayton	0.978	0.014	-6936
8 hours	9 767	t	0.619; 4.833	0.007; 0.293	-5006
		Gumbel	1.688	0.014	-4561
		Gaussian	0.610	0.005	-4540
		Frank	4.633	0.073	-4347
		Clayton	1.020	0.020	-3697
12 hours	6 513	t	0.623; 5.438	0.008; 0.449	-3350
		Gaussian	0.617	0.007	-3111
		Gumbel	1.689	0.017	-3047
		Frank	4.680	0.089	-2953
		Clayton	1.037	0.025	-2518
1 day	3 259	t	0.624; 5.712	0.011; 0.714	-1675
		Gaussian	0.621	0.009	-1576
		Gumbel	1.689	0.024	-1525
		Frank	4.650	0.125	-1471
		Clayton	1.056	0.035	-1287

Table 2.1: *Parameter estimates, standard errors and Akaike's information criterion values for the various copula models and time frequencies. For the t–copula the first parameter estimate is the correlation and the second is the degrees of freedom and respectively for the* $\hat{s.e.}$*'s.*

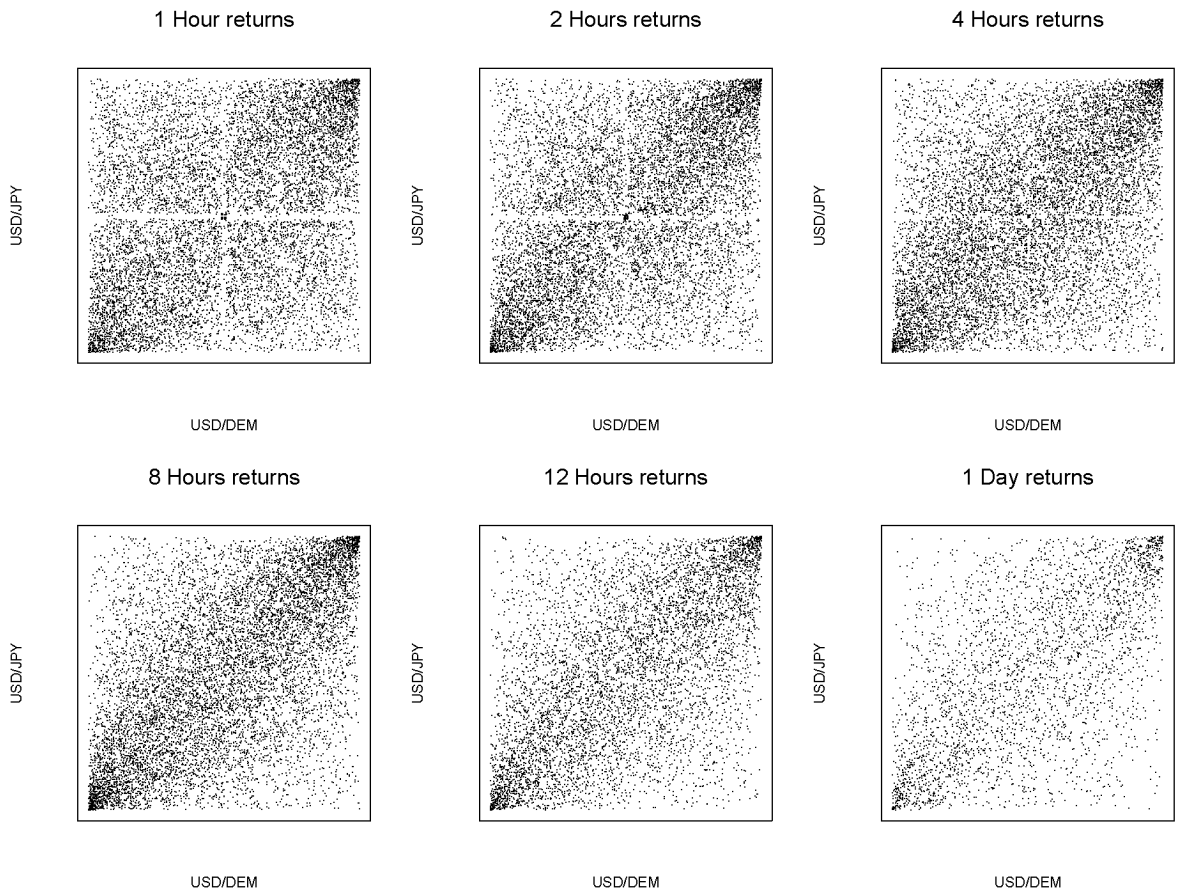

Figure 2.10: *Bivariate returns of USD/DEM and USD/JPY for different time frequencies mapped into the unit square by its marginal empirical distributions.*

of some of these families are in Section 1.1. The remaining copulae and further details can be found in [28], [29], [48] and [63].

In Table 2.1 the parameter estimates, the corresponding standard errors, and the AIC values (1.26) are given for each of the models fitted. For every time frequency the models are ordered by their AIC value. The first observation is that, for the five models considered, the t–copula model has the best performance according to this criterion. In Figure 2.13 we plot, for each model and frequency, the AIC of the t–copula minus the AIC of the model and divided this difference by the number of observations (in order to give the plots a comparable scale). The degrees of freedom estimated for the t–copula are plotted in parentheses. We note that the degrees of freedom of the t–copula increase from $\hat{\nu} = 4.3$ for hourly returns

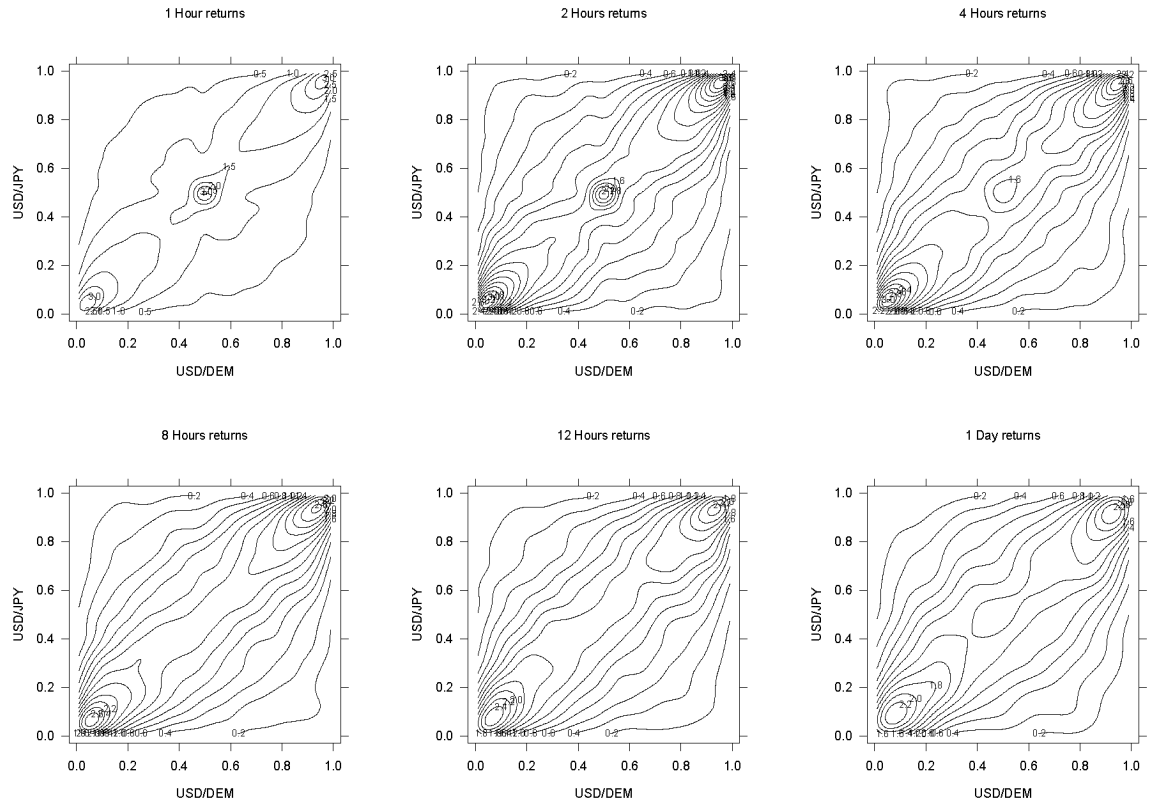

Figure 2.11: *Contour–plots of the bivariate returns of USD/DEM and USD/JPY for different time frequencies mapped into the unit square by its marginal empirical distributions.*

to $\hat{\nu} = 5.7$ for daily returns. This is similar to the behaviour of the tail index estimates for univariate data as a function of the time horizon; see Müller et al. [62]. It raises the question of what happens to the tail–dependence when the time frequency of the returns vary.

If we assume the t–copula as the model for the data, from the results in Table 2.1 we can then estimate the tail–dependence coefficient at the different time frequencies. This coefficient, defined in Proposition 1.14, was given for the t–copula in Example 1.15. Table 2.2 shows the tail coefficient estimates for the time frequencies considered for the bivariate USD/DEM and USD/JPY returns, assuming a t–model for the dependence structure. The values obtained indicate that the bivariate returns USD/DEM and USD/JPY remain asymptotically dependent across the time frequencies considered. A confidence interval analysis can be worked out like it was

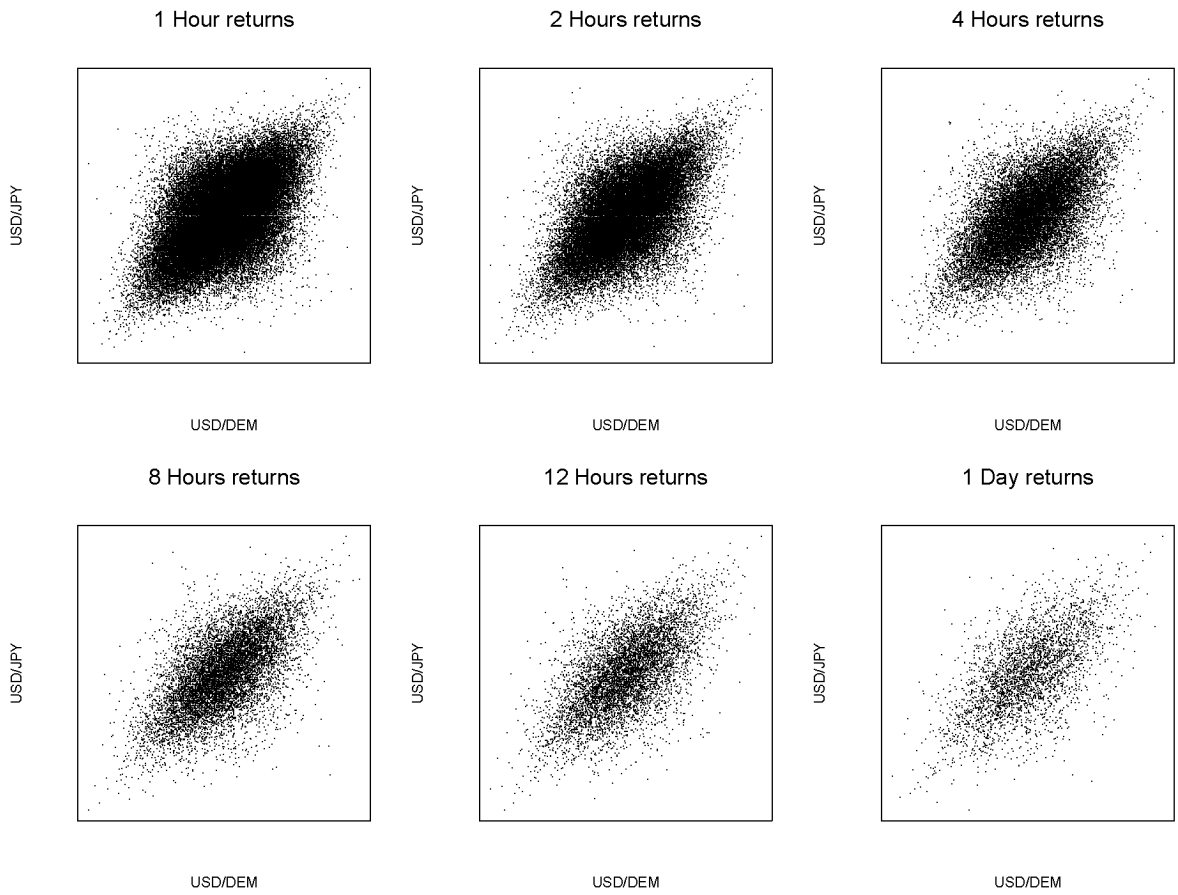

Figure 2.12: *Bivariate pseudo–returns for different time frequencies plotted with standard normal margins.*

done in Section 1.7.4 and as explained in Section 1.8.

2.3 Goodness–of–fit tests

In this section we test for the ellipticity of the data. Furthermore, we test for the goodness–of–fit of the copulae fitting using the test based on the probability–integral transformation introduced in Section 1.6.

Frequency	$\hat{\nu}$	$\hat{\rho}$	$\hat{\lambda}$
1 hour	4.339	0.563	0.273
2 hour	4.269	0.585	0.291
4 hour	4.282	0.599	0.299
8 hour	4.833	0.619	0.287
12 hour	5.438	0.623	0.264
1 day	5.712	0.624	0.254

Table 2.2: *Tail coefficient estimates for the USD/DEM and USD/JPY bivariate returns for the six different time frequencies considered.*

2.3.1 Test of elliptical symmetry

The test of elliptical symmetry used here is due to Manzotti et al. [59]. Suppose that $\mathbf{X}$ is a d–dimensional random vector with an elliptical distribution. $\mathbf{X}$ can then be represented as

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{R}A\mathbf{U},$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$, A is a non-singular $d \times d$ matrix, $\mathbf{R}$ is a real non negative random variable, $\mathbf{U}$ is uniformly distributed on the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \| \mathbf{x} \| = 1\}$ and $\mathbf{R}$ and $\mathbf{U}$ are independent. Let $\Sigma = AA'$ be the shape matrix. The covariance matrix of $\mathbf{X}$, Σ_0 is proportional to Σ. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be an iid sample from a d–dimensional distribution. The null hypothesis of the test is that the sample comes from an elliptically distributed population. Let $\overline{\mathbf{X}}$ and S denote the sample mean and covariance matrix, respectively. Consider the scaled residuals $\mathbf{Y}_k = S^{-1/2}(\mathbf{X}_k - \overline{\mathbf{X}})$ for $k = 1, \ldots, n$. Let $\mathbf{W}_k = \mathbf{Y}_k / \| \mathbf{Y}_k \|$ for $k = 1, \ldots, n$ be the projections of the scaled residuals on the unit sphere. If $\mathbf{X}$ is elliptically symmetric then $\mathbf{W}$ is approximately uniformly distributed on $\mathbb{S}^{d-1}$.

Consider $\varepsilon > 0$ fixed and let n_ε be the integer part of εn. Let q_n be the ε empirical quantile for the variables $\| \mathbf{Y}_1 \|, \| \mathbf{Y}_2 \|, \ldots, \| \mathbf{Y}_n \|$. With the average

$$Q_n(h) = \frac{1}{n} \sum_{k=1}^{n} h(\mathbf{W}_k) \mathbb{I}_{\{\| \mathbf{Y}_k \| > q_n\}},$$

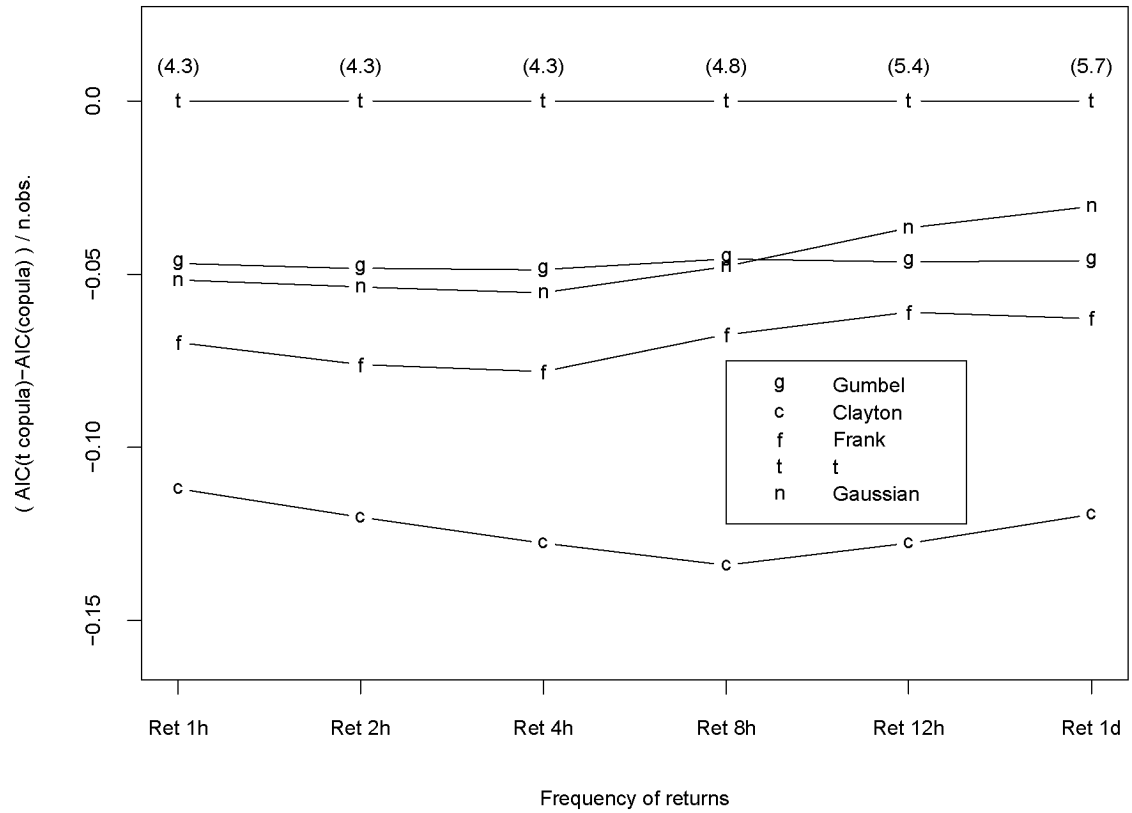

Figure 2.13: *Comparison of the AIC values for the different frequencies.*

where h is a function defined on $\mathbb{S}^{d-1}$, the test statistic Z_n^2 is given by

$$Z_n^2 = n \sum_{h \in \mathcal{J}_{jl}} Q_n^2(h)$$

for $j \geq 3$. The set of functions $\mathcal{J}_{jl}$ is the union $\mathcal{J}_{jl} = \bigcup_{j \leq i \leq l} \mathcal{H}_i$, where $\mathcal{H}_j$ denotes the set of spherical harmonics of degree j in the orthonormal basis as considered in [59]. This test statistic consists on averaging spherical harmonics over the projections of the $\mathbf{Y}_k$'s on the unit sphere. It will be useful to know that there are $N(d,j) = \binom{d+j-1}{j} - \binom{d+j-3}{j-2}$ linearly independent spherical harmonics of degree j in dimension d. Let N denote the number of functions in $\mathcal{J}_{jl}$. The main result of Manzotti et al. [59] states that, independently of the unknown parameters of the distribution under the null hypothesis, the limiting distribution of Z_n^2 is that of $(1 -$

$\varepsilon)\chi^2$, where χ^2 has a chi–square distribution with N degrees of freedom.

With the $d = 2$ dimensional data we used $\varepsilon = 0.05$, $j = 3$ and $l = 6$, so that $N = 8$ and the orthonormal spherical harmonics used are

$$h_{1,j}(\mathbf{W}_k) = 2^{1/2}\cos(j\theta_k),\ h_{2,j}(\mathbf{W}_k) = 2^{1/2}\sin(j\theta_k),$$

for $3 \leq j \leq 6$ and $\mathbf{W}_k = (\cos(\theta_k), \sin(\theta_k))$.

The estimators

The estimation of the sample covariance matrix requires some care because in this elliptical setting, we often have heavy tailed margins. In this case the standard estimators may have a poor performance. In order to test for ellipticity it is enough to estimate a matrix which is proportional to the covariance matrix. Indeed, if $\mathbf{X} = \boldsymbol{\mu} + \mathbf{R}A\mathbf{U}$ is elliptical then $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{R}(cA)\mathbf{U}$, for $c > 0$, is also elliptical. Having this in mind, in the bivariate case, we can estimate the matrix

$$\sigma_1^{-2}\Sigma = \begin{pmatrix} 1 & \rho\frac{\sigma_2}{\sigma_1} \\ \rho\frac{\sigma_2}{\sigma_1} & \frac{\sigma_2^2}{\sigma_1^2} \end{pmatrix}$$

using more robust estimators. Here, σ_i^2 are the diagonal elements of Σ. For the linear correlation coefficient, under the elliptical assumption, the estimator based on Kendall's tau,

$$\hat{\rho}_\tau = \sin\left(\frac{\pi}{2}\hat{\tau}\right),$$

is more efficient and robust than Pearson's linear correlation estimator; see Lindskog et al. [56]. In order to estimate the ratio between the standard deviations we can use another dispersion estimator rather than the standard, $S_n^2 = \sum_{i=1}^n (x_i - \overline{x})^2/n$, for instance the median absolute deviation, mad_n. The latter is more robust against extreme observations coming from heavy tails; see for instance Hampel et al. [43] or Pham–Gia and Hung [68]. The sample median absolute deviation is defined as follows:

$$mad_n(\mathbf{x}) = median(|\mathbf{x} - median(\mathbf{x})|),$$

where $median(\mathbf{x})$ denotes the sample median of the vector of observations $\mathbf{x}$. We then use

$$\widehat{\sigma_2/\sigma_1} = mad_n(\mathbf{x_2})/mad_n(\mathbf{x_1})$$

as an estimator more robust than the ratio of sample standard deviations.

2.3.2 Testing the results of the fittings

In Table 2.1 we ranked the different models fitted according to their AIC values. Table 2.3 contains the p–values for the probability integral transformation goodness–of–fit test explained in Section 1.6. We have listed only the results for the best fitting models (minimal AIC) from Table 2.1. In the same table are also the p–values (fourth column) for the ellipticity test described in Section 2.3.1. To test whether the values given by (1.24) come from a χ^2 distribution we use the Anderson–Darling goodness–of–fit test (see [1]). Looking at columns three and four it is only at a frequency of

Time frequency	Probability integral test		P–values for the ellipticity test	
	Model	P–value	Original margins	t margins
1 hour	t	0	0	0
2 hours	t	0	0	0
4 hours	t	0.01	0	0.092
8 hours	t	0.27	0	0.231
12 hours	t	0.19	0.034	0.369
1 day	t	0.74	0.821	0.675

Table 2.3: *P–values of goodness–of–fit and ellipticity tests.*

one day that we cannot reject ellipticity and have strong support for a t–copula. At higher frequencies, the situation is more subtle. The t–copula fits well up to eight hours. For four hours and higher, a more careful analysis (and possibly a more intricate copula) is needed, especially as at those frequencies very large sample sizes (e.g. 77 758 bivariate hourly observations) are available. Note that the null hypothesis of ellipticity is rejected for frequencies higher than one day. In column five of Table 2.3, we perform the same ellipticity test, but now after transforming the margins to a t–distribution. With the degrees of freedom estimated for the different frequencies, given in the Table 2.1, we compute the transformed

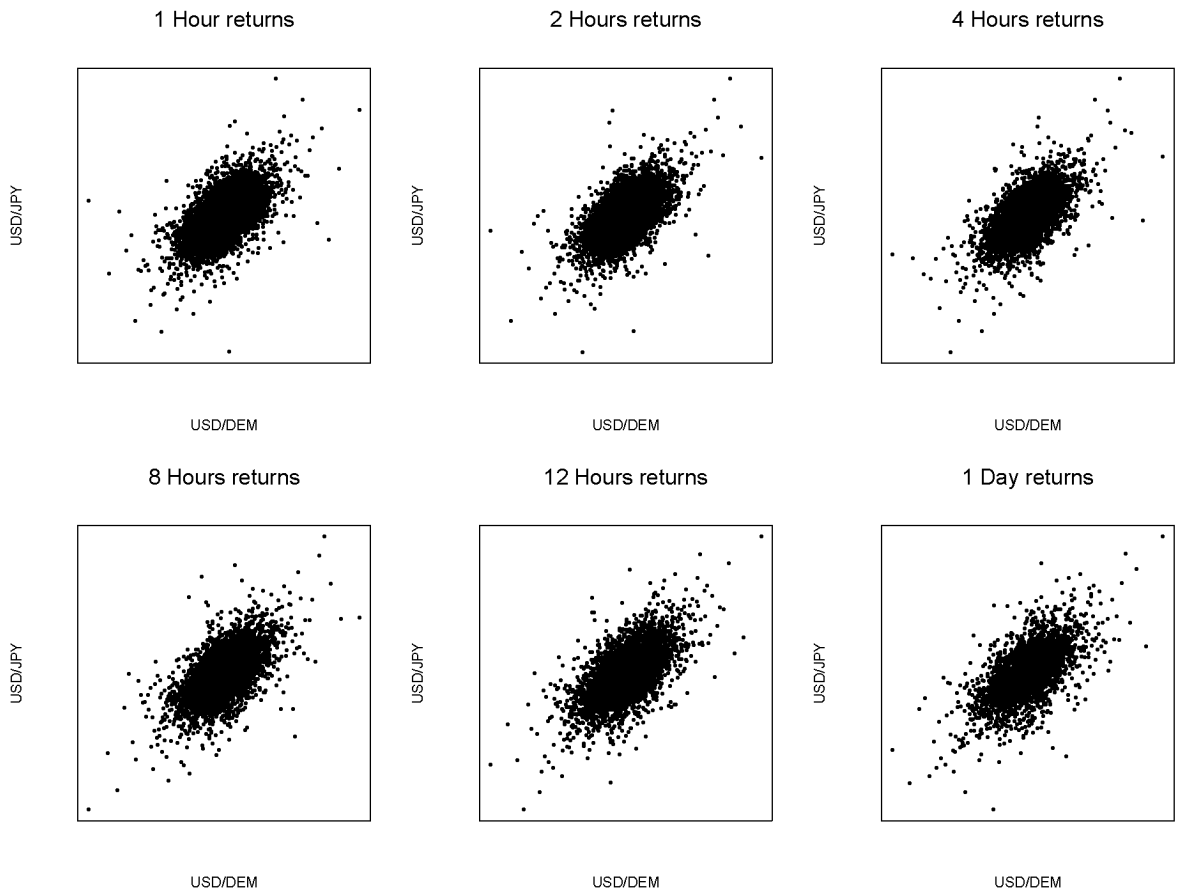

Figure 2.14: *Bivariate pseudo–returns for different time frequencies plotted with t margins.*

observations

$$\left(t_{\hat{\nu}}^{-1}\left(F_{1n}\left(x_{1i}\right)\right), t_{\hat{\nu}}^{-1}\left(F_{2n}\left(x_{2i}\right)\right)\right) \tag{2.10}$$

for $i = 1, 2, \ldots, n$ and for the six frequencies. As usual, $t_{\hat{\nu}}^{-1}$ denotes the quantile function of a standard univariate t–distribution with $\hat{\nu}$ degrees of freedom. In this way we try to reduce the chances that a rejection of the elliptical structure could come from non–elliptical margins. These new data are plotted in Figure 2.14. Ellipticity is now rejected only at the one and two hour frequency.

We continue our analysis by looking at the important problem of clustering between extreme movements, in the literature also referred to as tail–dependence.

2.4 Tail–dependence

Several authors have looked at the issue of tail–dependence in financial return data. See for instance Stărică [79] for a more mathematical discussion. Several references in Embrechts [25] yield guidance towards the more economic oriented literature. By definition, the notion of tail–dependence concerns bivariate rare events, hence typically, probabilistic limit theorems lie at the basis of any analysis. In order to investigate these problems on real data, a large number of observations is desirable. In Section 2.4.1 we start the study of the bivariate tail–dependence with an estimation of the spectral measure for all the time horizons considered in the previous sections. We then concentrate in Section 2.4.2 on hourly data only. Bivariate extremes in these data will be analysed using the theory of multivariate regular variation, leading to a spectral analysis, and a statistical analysis of bivariate excesses over high thresholds.

2.4.1 Spectral measure estimation

The mathematics underlying this section is to be found in Resnick [71] and Stărică [79]. Below we highlight the main definitions and notation. Let $\| \cdot \|$ denote the usual Euclidean L_2 norm on $\mathbb{R}^d$ and $\mathbb{S}^{d-1}$ be the unit sphere,

$$\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \| \mathbf{x} \| = 1\}.$$

Suppose that $\mathbf{X}$ is a random vector with values in $\mathbb{R}^d$. $\mathbf{X}$ is said to be multivariate regularly varying with tail index $\alpha > 0$ if there exists a random vector Θ such that for all $x > 0$, as $t \to \infty$,

$$\frac{P(\| \mathbf{X} \| > tx, \mathbf{X}/\| \mathbf{X} \| \in \cdot)}{P(\| \mathbf{X} \| > t)} \quad \xrightarrow{v} \quad x^{-\alpha} P(\Theta \in \cdot), \tag{2.11}$$

where $\xrightarrow{v}$ denotes vague convergence and Θ is a random vector on the space $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$. The distribution of Θ is referred to as the spectral distribution of $\mathbf{X}$. Definition (2.11) is equivalent to the existence of a

measure ν and a positive sequence (a_n), $a_n \to \infty$, such that for $n \to \infty$,

$$nP(a_n^{-1}\mathbf{X} \in \cdot) \quad \xrightarrow{v} \quad \nu(\cdot). \tag{2.12}$$

For a more precise and detailed treatment on this see for instance Resnick [70]. The measure ν has the following scaling property:

$$\nu(vS) = v^{-\alpha}\nu(S), \tag{2.13}$$

for any Borel set $S \subset [-\infty, \infty]^d \backslash \{0\}$. This property will be useful in order to find an estimator for the spectral distribution. Intuitively, α indicates the heaviness of the multivariate tails whereas $P(\Theta \in \cdot)$ measures in which parts of the space extremes cluster.

Define for $\mathbf{x} \in \mathbb{R}^d$ and $B \in \mathcal{B}(\mathbb{R}^d)$

$$\epsilon_x(B) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{if } x \in B^c. \end{cases}$$

Then a consistent estimator of $c\nu$, for some $c > 0$, is given by

$$\nu_n := \frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{\mathbf{X}_i / b(n/k_n)},$$

where $b(.)$ is the quantile function $b(t) := F^{\leftarrow}(1 - 1/t)$, for $t > 1$, of the random variable $\| \mathbf{X} \|$. As usual in extreme value theory, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$; see Resnick [71]. If we estimate the quantile function with the corresponding empirical estimator

$$\hat{b}\left(\frac{n}{k_n}\right) = \| \mathbf{X} \|_{k_n,n},$$

where $\| \mathbf{X} \|_{k_n,n}$ is the k_nth largest value of the one–dimensional set $\{\| \mathbf{X}_i \|: i = 1, 2, \ldots, n\}$, we obtain as estimator of the spectral distribution

$$\hat{\nu}_n = \frac{1}{k_n} \sum_{i=1}^{n} \epsilon_{\mathbf{x}_i / \|\mathbf{x}\|_{k_n,n}}(V(S)) \tag{2.14}$$

where $V(S) = \{\mathbf{x} \in \mathbb{S}_+^{d-1} : \mathbf{x}/ \| \mathbf{x} \| \in S\}$ and $\mathbb{S}_+^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \| \mathbf{x} \| > 1\}$. The performance of this estimator very much depends on the choice of

k_n. Here we use the scaling property (2.13) and choose k_n such that $\hat{\nu}_n(u\mathbb{S}_+^{d-1})/(u^{-\alpha_{k_n}}\hat{\nu}_n(\mathbb{S}_+^{d-1})) \approx 1$ for values of u in a neighbourhood of 1. We plot the set of values

$$\left\{\left(u, \frac{\hat{\nu}_n(u\mathbb{S}_+^{d-1})}{u^{-\hat{\alpha}_{k_n}}\hat{\nu}_n(\mathbb{S}_+^{d-1})}\right) : 0 < u < 2\right\} \tag{2.15}$$

for several values of k_n and choose the one corresponding to the plot for which these values are closer to 1 around $u = 1$. These are the so–called Stărică plots. For more on this see Stărică [79]. These plots are given in Figure 2.15 for the six bivariate pseudo–return series. The number of excesses k_n and the the tail index estimates used in (2.15) to produce the plots in Figure 2.15 are in Table 2.4. In order to obtain the tail index estimate $\hat{\alpha}_{k_n}$, we use the Hill estimator,

$$\hat{\alpha}_{k_n} = \left(\frac{1}{k_n}\sum_{j=1}^{k_n}\log x_{j,n} - \log x_{k_n,n}\right)^{-1}, \tag{2.16}$$

where theoretically $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$ so as to ensure weak consistency (see Embrechts et al. [27, page 336]). Standard errors can be calculated; we do not report them as for the analysis of extremal clustering we only need the point estimates.

Time frequency	Sample size n	Number of excesses k_n	Tail index estimate $\hat{\alpha}_{k_n}$
1 hour	77 758	500	4.085
2 hours	38 976	350	3.943
4 hours	19 514	370	3.747
8 hours	9 767	300	4.163
12 hours	6 513	350	3.941
1 day	3 259	260	3.931

Table 2.4: *Number of excesses and Hill estimates for the tail index of the bivariate tail returns on USD/DEM and USD/JPY spot rates.*

We now estimate the spectral density of the bivariate returns $\mathbf{X}$ of the USD/DEM and USD/JPY data at a given time frequency using (2.14).

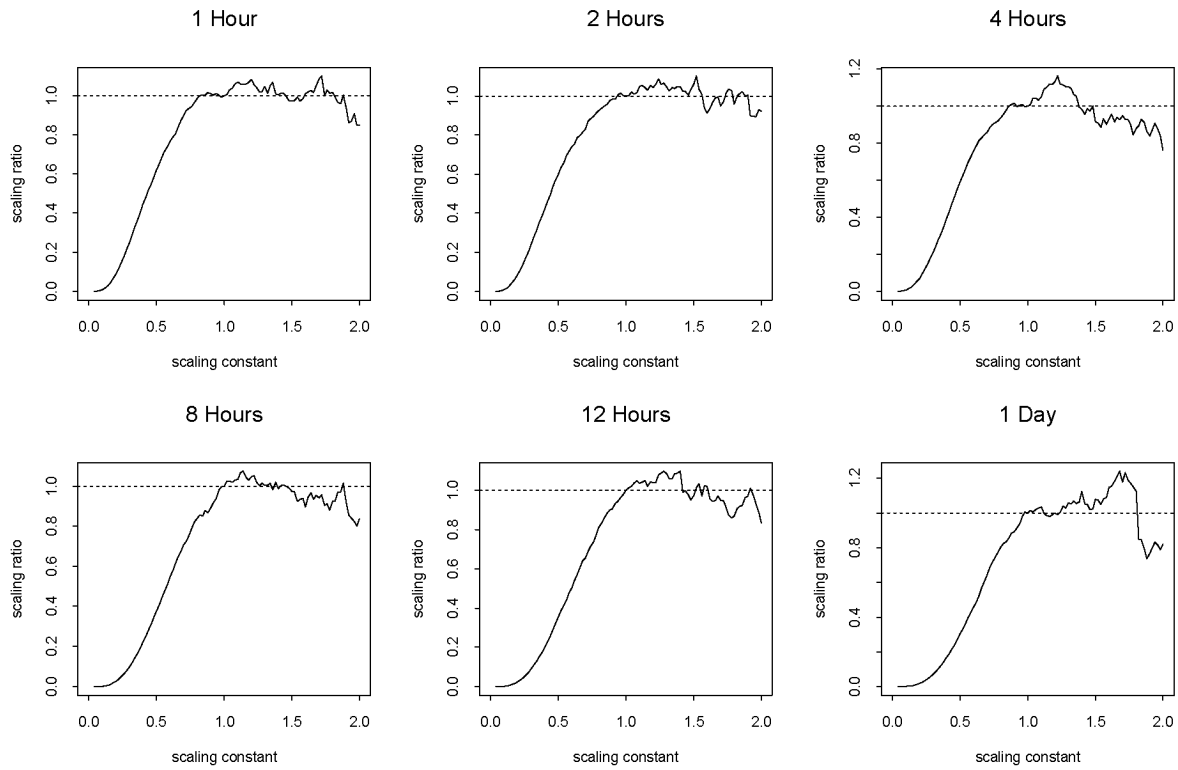

Figure 2.15: *Stărică plots for the six bivariate pseudo–return series. The number of excesses is* $k_{77\,758} = 500$, $k_{38\,976} = 350$, $k_{19\,514} = 370$, $k_{9\,767} = 300$, $k_{6\,513} = 350$ *and* $k_{3\,259} = 260$ *respectively from one hour up to one day frequencies. The scaling constant is* u *in (2.15) and the scaling ratio is* $\hat{\nu}_n(u\mathbb{S}_+^{d-1})/(u^{-\hat{\alpha}_{k_n}}\hat{\nu}_n(\mathbb{S}_+^{d-1}))$.

First, choose k_n as described above and consider the points

$$\left\{\theta_i \in [0, 2\pi[: (\cos\theta_i, \sin\theta_i) = \frac{\mathbf{x}_i}{\| \mathbf{x}_i \|}, \ \| \mathbf{x}_i \| > \| \mathbf{x} \|_{k_n,n}, i = 1, \ldots, n\right\}. \tag{2.17}$$

We then plot a non–parametric density estimate for these angular observations using a smoothed kernel estimator with Gaussian weights and bandwidth 0.2π. In a more detailed analysis one could also work out confidence bands around the estimated functions. In Figure 2.16 we have plotted the estimated spectral densities for each time frequency. We would first like to point out that a regularly varying spherical distribution has a uniform spectral distribution when the Euclidean norm is used in its definition. In this case, a fairly constant estimated spectral density would appear. Peakedness in one or other direction points at clustering of ex-

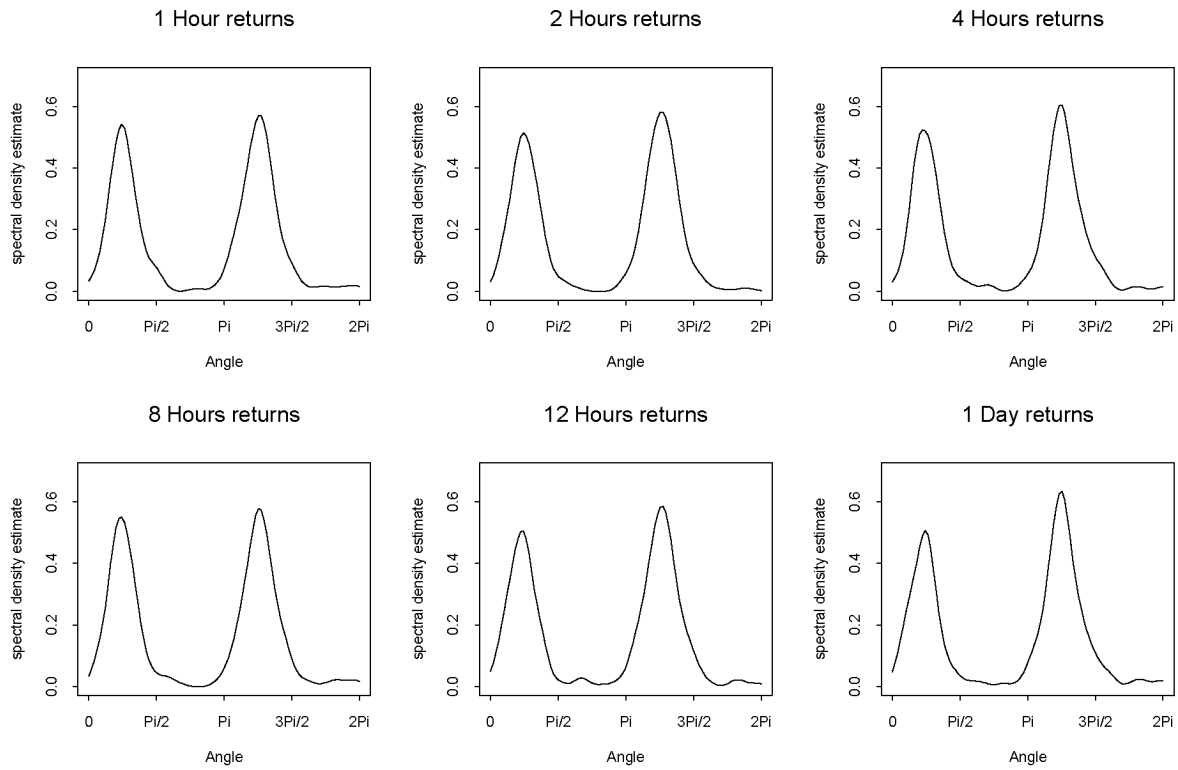

Figure 2.16: *The estimated spectral densities for the different time frequencies.*

tremes in that direction. Also, the procedure discussed only uses data in the tails so that no information from the centre of the distribution enters. Figure 2.16 clearly shows clustering of extreme returns in the first and third quadrants, also referred to as positive dependence. This dependence persists at all frequencies and turns out to be fairly symmetric. Of course, a basic assumption concerns the property (2.11); as in the one–dimensional case, one can show, using extreme value theory, that the limit property (2.11) is very natural for multivariate financial return data. For an interesting paper leading to similar conclusions, see Hauksson et al. [45].

2.4.2 Multivariate excesses

In the previous section, under the assumption (2.11), we modelled the occurrence of joint extremes for bivariate returns by the density of the

spectral random variable Θ. In this section we focus, for a given high threshold t, on the event $\{X_1 > t, X_2 > t\}$. Similarly we can define sets of returns simultaneously smaller than a given threshold. As the threshold t will be large (small) we will concentrate only on the one hour frequency which allows for sufficient data. In Figure 2.17 we plotted the bivariate excesses for different values of the threshold. For univariate random variables, the Balkema–de Haan–Pickands result (Theorem 3.4.13(b) in Embrechts et al. [27]) yields the generalised Pareto distribution as a canonical model for the distribution function of conditional excesses. In Juri and Wüthrich [49], a similar result is proved for bivariate excesses in the case of Archimedean copulae (note that in that case, one makes an assumption on the copula for the whole domain of the bivariate dependence structure model). We summarise below this result which forms the basis for our statistical analysis. The copula $C(u, v)$, with $0 \leq u, v \leq 1$, of the random vector (U, V) is called Archimedean if there exists a continuous, strictly decreasing function, $\psi : [0, 1] \mapsto [0, \infty]$ with $\psi(1) = 0$, such that

$$C(u, v) = \psi^{[-1]}(\psi(u) + \psi(v)). \tag{2.18}$$

The function $\psi^{[-1]} : [0, \infty] \mapsto [0, 1]$ is defined by

$$\psi^{[-1]}(x) = \psi^{-1}(x)\mathbb{I}_{[0,\psi(0)]}(x)$$

and is called the generator of the copula C. Denote by F_t the conditional distribution function

$$F_t(u) := P(U \leq u | U \leq t, V \leq t),\ 0 \leq u \leq 1.$$

The extreme tail–dependence copula of the copula C relative to a threshold t is given by

$$C_t(u, v) = P(U \leq F_t^{-1}(u), V \leq F_t^{-1}(v) | U \leq t, V \leq t).$$

If C is an Archimedean copula having a regularly varying differentiable generator $\psi \in \mathcal{R}_{-\alpha}$ with $0 < \alpha < \infty$, then

$$\lim_{t \to 0^+} C_t(u, v) = C_\alpha^{Cl}(u, v), \tag{2.19}$$

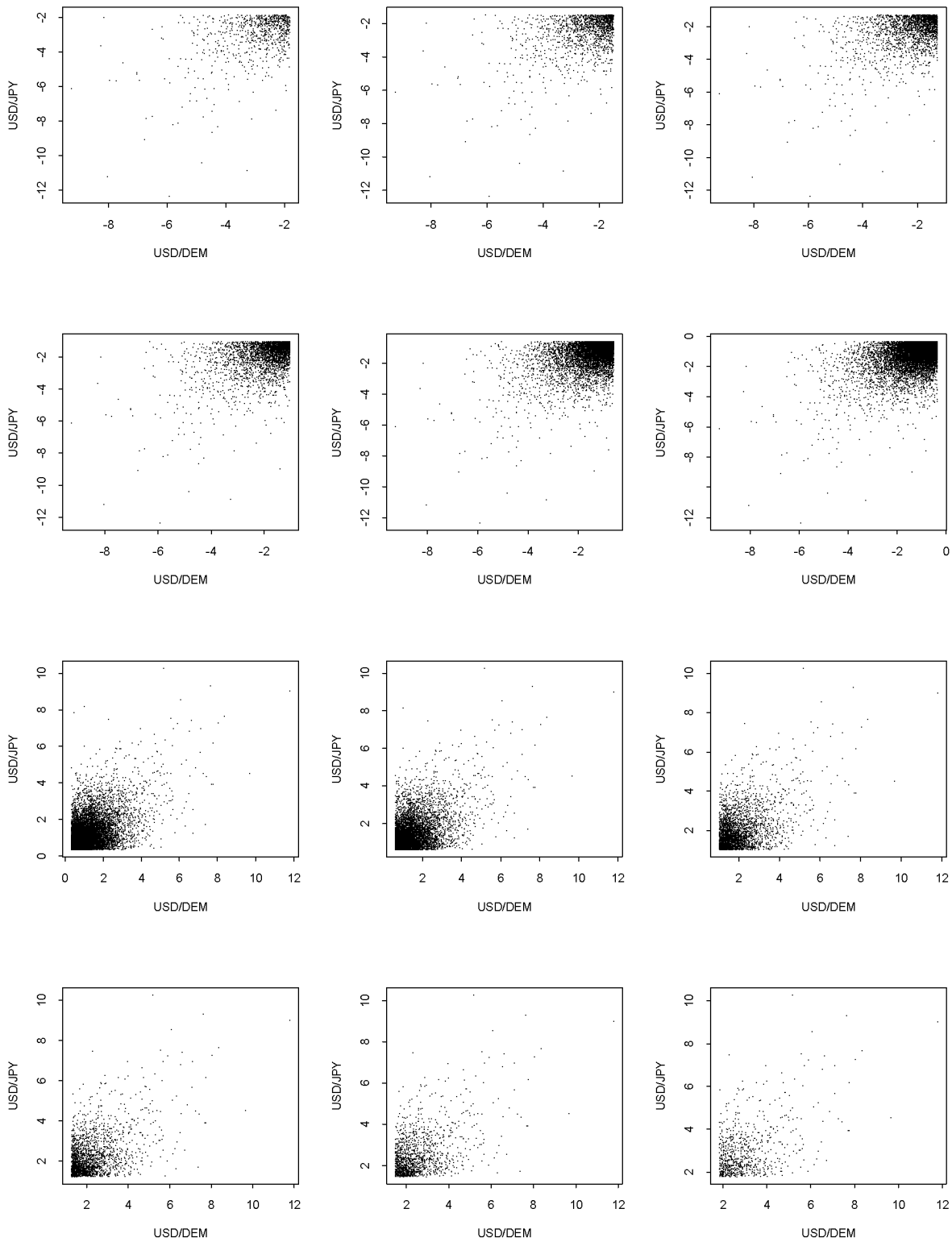

Figure 2.17: *The bivariate excesses of the one hour returns for different thresholds.*

for all $0 \leq u, v \leq 1$. The limit copula in this result, C_α^{Cl}, is the Clayton copula with parameter $\alpha > 0$ defined by

$$C_\alpha^{Cl}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha},$$

for $0 < u\,, v \leq 1$. Juri and Wüthrich [49] show that the condition $\psi \in \mathcal{R}_{-\alpha}$ is a very natural one which holds for several known examples.

For the one hour pseudo–returns of USD/DEM and USD/JPY, $(F_{1n}(x_{1i}), F_{2n}(x_{2i}))$ with $i = 1, \ldots, n$, we considered several thresholds t (both in the joint left as well as in the joint right tails) and fitted copula models C_t. The thresholds and the resulting data are to be found in Figure 2.18. Hence we want to model

$$C_{t-}(u, v) = P(U \leq F_t^{-1}(u), V \leq F_t^{-1}(v) | U \leq t, V \leq t) \tag{2.20}$$

as well as

$$C_{t+}(u, v) = P(U \leq F_t^{-1}(u), V \leq F_t^{-1}(v) | U \geq t, V \geq t). \tag{2.21}$$

In each case we fitted a list of copula models including the Gaussian, t, Gumbel, Frank, Clayton, survival Gumbel, survival Clayton and Farlie–Gumbel–Morgenstern copulae. For the definition of the copulae used, see again Section 1.1 and [29], [48] and [63]. The survival copula was introduced in Proposition 1.8. The results are reported in Tables 2.5 and 2.6 where the models fitted are ranked by Akaike's information criterion. In Table 2.5 are the results for C_{t-}. The second column contains the number of observations below t in percentage of the total data. For the t–copula, the first parameter estimate is the correlation and the second is the degrees of freedom. The Clayton copula gives always the best fitting. Remember that the best fitting for the dependence structure of the full hourly data set was attained with the t–copula, although without passing the goodness–of–fit test. Note that, for the considered thresholds, the Clayton parameter ranges from $\hat{\alpha} = 0.556$ to $\hat{\alpha} = 0.609$ which corresponds to a Kendall tau coefficient between $\hat{\tau} = 0.217$ and $\hat{\tau} = 0.233$ (for the Clayton copula with parameter α, $\tau = \alpha/(\alpha + 2)$). Table 2.6 contains the

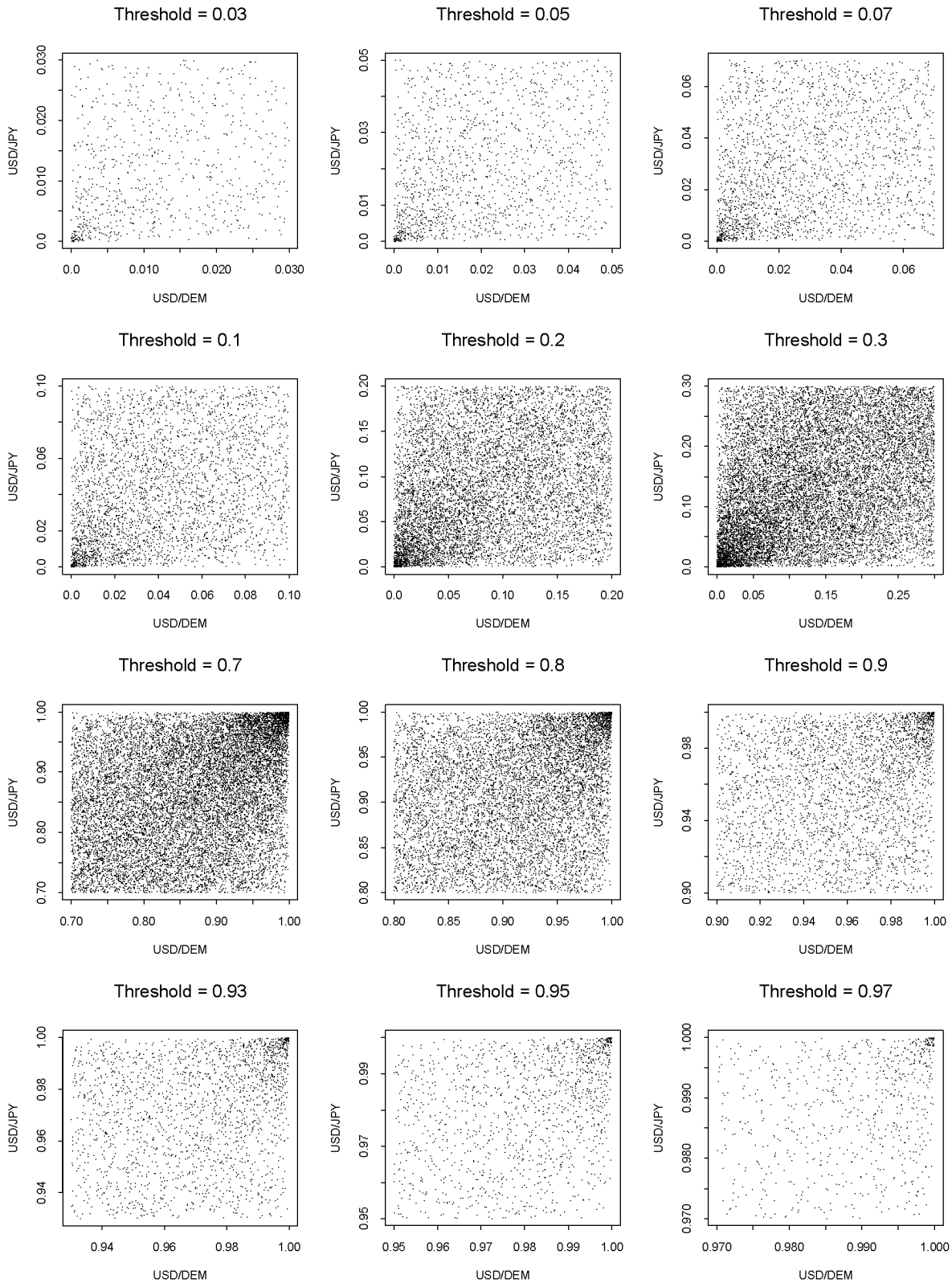

Figure 2.18: *The bivariate excesses of the one hour returns mapped into the unit square by the empirical distributions of the margins. Only the tail regions are shown.*

results for C_{t+}. In this case, the survival Clayton copula yields the best fit. In this case the survival Clayton parameter varies between $\hat{\alpha} = 0.574$ and $\hat{\alpha} = 0.666$, corresponding to a Kendall tau from $\hat{\tau} = 0.223$ to $\hat{\tau} = 0.250$. Both tables are summarised graphically in Figure 2.19 (similar to Figure 2.13 for the full copula data).

The threshold considered in this section for the modelling of the conditional copula is the same for both margins. In Charpentier [11] a generalisation of result (2.19) is given where a different threshold for each margin can be used. The Clayton copula remains the limiting conditional copula in this generalisation.

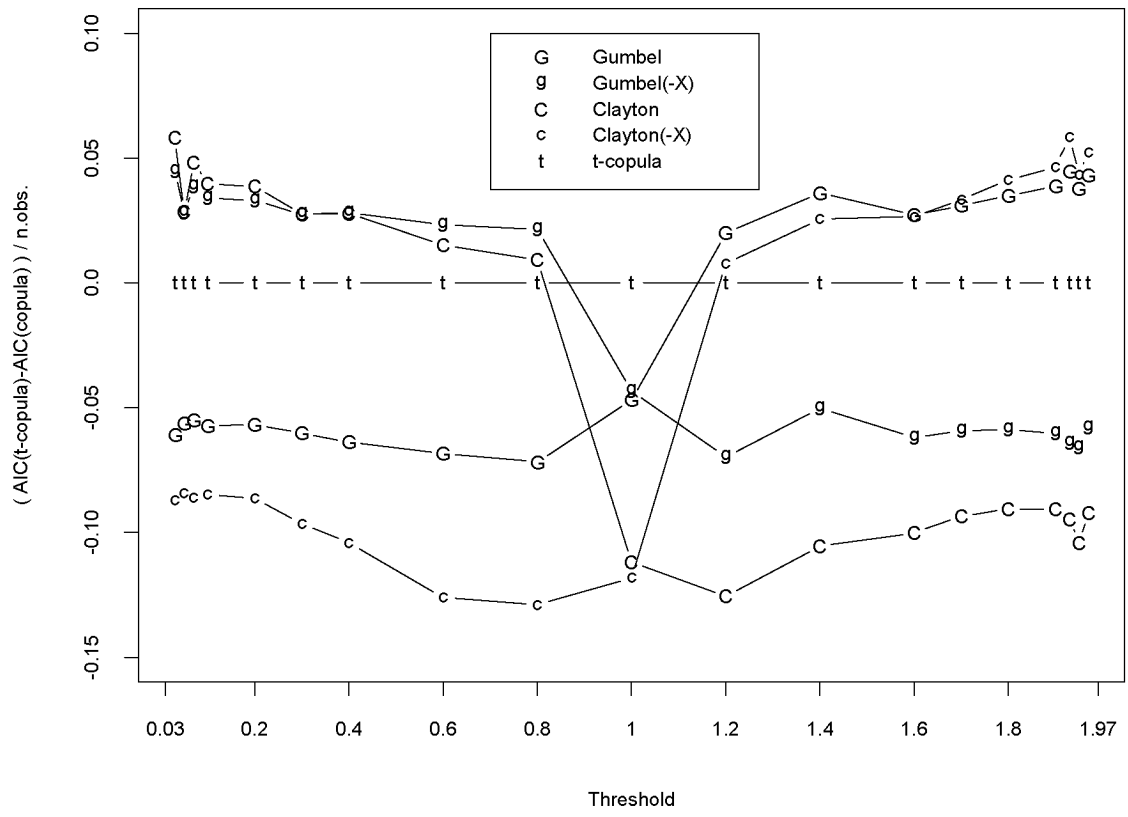

Figure 2.19: *Comparison of the AIC values for the different thresholds.*

Threshold	n. obs.	Model	Estimate	$s\hat{e}$.	AIC	p–value
		Clayton	0.583	0.059	-132.19	0.99
		Surv. Gumbel	1.295	0.034	-122.04	0.99
		t	0.335; 13.69	0.033; 8.219	-88.00	0.99
0.03^-	759	Gaussian	0.336	0.030	-86.79	0.92
	(0.98%)	Frank	2.031	0.227	-78.13	0.95
		F–G–M	0.873	0.086	-70.48	0.40
		Gumbel	1.195	0.033	-41.71	0.89
		Surv. Clayton	0.238	0.052	-21.81	0.45
		Clayton	0.561	0.043	-227.74	0.98
		Surv. Gumbel	1.298	0.025	-227.13	0.99
		t	0.342; 8.40	0.025; 2.31	-188.48	0.97
0.05^-	1 376	Gaussian	0.348	0.022	-173.46	0.68
	(1.77%)	Frank	2.070	0.169	-147.70	0.78
		F–G–M	0.863	0.062	-129.49	0.10
		Gumbel	1.223	0.024	-110.96	0.78
		Surv. Clayton	0.305	0.039	-72.38	0.23
		Clayton	0.556	0.035	-350.71	0.99
		Surv. Gumbel	1.283	0.020	-330.20	0.93
		t	0.330; 11.44	0.019; 3.35	-248.47	0.95
0.07^-	2 112	Gaussian	0.328	0.018	-236.91	0.79
	(2.72%)	Frank	1.999	0.135	-216.32	0.84
		F–G–M	0.903	0.054	-204.21	0.17
		Gumbel	1.200	0.019	-132.03	0.86
		Surv. Clayton	0.244	0.031	-66.80	0.27
		Clayton	0.558	0.028	-547.96	0.91
		Surv. Gumbel	1.289	0.016	-529.62	0.61
		t	0.340; 13.43	0.015; 3.62	-418.10	0.85
0.1^-	3 273	Gaussian	0.342	0.014	-403.71	0.81
	(4.21%)	Frank	2.046	0.108	-352.09	0.74
		F–G–M	0.897	0.040	-326.39	0.07
		Gumbel	1.209	0.015	-230.55	0.81
		Surv. Clayton	0.273	0.025	-140.27	0.15
		Clayton	0.556	0.018	-1302.15	0.96
		Surv. Gumbel	1.289	0.010	-1257.91	0.79
		t	0.340; 12.19	0.010; 1.90	-1000.15	0.82
0.2^-	7 807	Gaussian	0.339	0.009	-952.77	0.21
	(10.0%)	Frank	2.085	0.070	-870.88	0.47
		F–G–M	0.905	0.026	-803.05	0.001
		Gumbel	1.211	0.010	-555.75	0.50
		Surv. Clayton	0.271	0.016	-324.74	0.01
		Clayton	0.609	0.014	-2547.77	0.47
		Surv. Gumbel	1.330	0.008	-2546.12	0.68
		t	0.383; 12.59	0.007; 1.52	-2178.67	0.85
0.3^-	13 359	Gaussian	0.381	0.006	-2097.90	0.06
	(17.2%)	Frank	2.401	0.054	-1944.88	0.39
		F–G–M	0.982	0.015	-1750.99	0.00
		Gumbel	1.260	0.008	-1374.71	0.20
		Surv. Clayton	0.348	0.012	-889.42	0.00

Table 2.5: *Fitting results for bivariate excesses on the third quadrant of one hour returns for different thresholds.*

Threshold	n. obs.	Model	Estimate	$\hat{s.e.}$	AIC	p–value
		Surv. Clayton	0.601	0.060	-137.23	0.94
		Gumbel	1.306	0.035	-130.25	0.97
		t	0.343; 8.42	0.034; 3.45	-98.15	0.98
0.97^+	745	Gaussian	0.350	0.030	-93.00	0.91
	(0.96%)	Frank	2.048	0.229	-77.25	0.96
		F–G–M	0.878	0.090	-68.27	0.33
		Surv. Gumbel	1.221	0.033	-54.98	0.93
		Clayton	0.278	0.054	-29.41	0.42
		Surv. Clayton	0.666	0.046	-288.12	0.92
		Gumbel	1.351	0.027	-280.52	0.57
		Gaussian	0.403	0.021	-230.68	0.94
0.95^+	1 331	t	0.400; 23.09	0.022; 16.97	-230.50	0.84
	(1.71%)	Frank	2.417	0.172	-196.73	0.71
		F–G–M	0.999	0.072	-189.17	0.25
		Surv. Gumbel	1.267	0.026	-142.60	0.87
		Clayton	0.359	0.041	-91.74	0.52
		Surv. Clayton	0.597	0.036	-378.41	0.99
		Gumbel	1.299	0.021	-350.57	0.92
		t	0.339; 11.25	0.020; 3.40	-260.50	0.95
0.93^+	2 014	Gaussian	0.344	0.018	-249.74	0.92
	(2.59%)	Frank	2.032	0.139	-212.26	0.86
		F–G–M	0.896	0.054	-196.21	0.15
		Surv. Gumbel	1.203	0.020	-131.48	0.93
		Clayton	0.248	0.032	-69.92	0.30
		Surv. Clayton	0.583	0.028	-575.19	0.93
		Gumbel	1.299	0.017	-550.75	0.78
		t	0.345; 11.97	0.016; 3.13	-428.26	0.89
0.9^+	3 167	Gaussian	0.351	0.014	-413.18	0.82
	(4.07%)	Frank	2.050	0.110	-340.33	0.83
		F–G–M	0.896	0.043	-309.91	0.05
		Surv. Gumbel	1.215	0.016	-236.88	0.80
		Clayton	0.277	0.025	-141.00	0.10
		Surv. Clayton	0.574	0.018	-1376.35	0.99
		Gumbel	1.298	0.010	-1327.02	0.83
		t	0.345; 10.76	0.010; 1.56	-1055.79	0.87
0.8^+	7 765	Gaussian	0.348	0.009	-1001.24	0.27
	(9.99%)	Frank	2.091	0.071	-868.78	0.41
		F–G–M	0.910	0.026	-797.18	0.00
		Surv. Gumbel	1.218	0.010	-599.50	0.41
		Clayton	0.280	0.016	-352.28	0.01
		Surv. Clayton	0.594	0.014	-2459.31	0.84
		Gumbel	1.315	0.008	-2426.10	0.60
		t	0.366;11.63	0.007; 1.36	-2014.98	0.75
0.7^+	13 300	Gaussian	0.367	0.007	-1928.50	0.09
	(17.1%)	Frank	2.253	0.054	-1718.03	0.27
		F–G–M	0.954	0.018	-1561.01	0.00
		Surv. Gumbel	1.242	0.008	-1226.49	0.22
		Clayton	0.320	0.012	-770.77	0.00

Table 2.6: *Fitting results for bivariate excesses on the first quadrant of one hour returns for different thresholds.*

Comments

In this chapter we analysed the dependence structure within two–dimensional, high–frequency FX return data. The methods used are copula modelling together with statistical techniques for extremal clustering. An overall picture emerged that is as follows: at all time horizons the data can be best fitted with t–copulae with successively higher degrees of freedom as the time horizon increases. Note that the t–copula is rejected for the shortest horizons probably because of the large amount of data. This means that the t–copula has not enough structure to properly describe the details which can be discerned within such a large sample. If the margins are transformed in order to be t–distributed, the test for ellipticity is not rejected except for the one hour and two hours horizons. With the original margins, ellipticity is rejected for horizons of eight hours and shorter. The spectral measure, however, shows pronounced peaks in the diagonals for all time horizons. An analysis of the multivariate excesses of hourly returns shows that the lower left tail is best described by Clayton or survival Gumbel copulae while the upper right tail is best described with Gumbel or survival Clayton copulae, as indeed predicted by theory.

These results extend the univariate stylised facts to the bivariate case and give valuable indications for time series models. Throughout, we used time–invariant stochastic copula–based models. As in the one–dimensional case, time series models allowing for stochastic volatility are to be analysed. The methods introduced are then to be used at the level of the residuals. This is the aim of Chapter 3.

Chapter 3

Conditional copula models

In the previous chapter we investigated the stylised facts of the dependence structure of the bivariate FX deseasonalised log–returns on the USD/DEM and USD/JPY at six different time horizons. It is important to stress that that analysis assumed the return vectors to be iid. We know however that this assumption is violated in practice due for instance to volatility effects. Recall the slow decay of the sample autocorrelation and cross–correlation functions in Figure 2.6. Hence, our analysis there concerned the stationary dependence structure, assumed to exist. In the present chapter we therefore start again from the deseasonalised FX data and investigate dependence between the (residual or filtered) vector components after some dynamic model has been fitted. We first filter the data through univariate ARMA–GARCH models and analyse the copula function of the residuals. Ellipticity is tested, spectral densities are estimated and the extreme tail–dependence copula is modelled.

3.1 Time dependence filtering

Among the empirical stylised facts for univariate financial returns are non–Gaussian innovations and conditional heteroscedasticity (ARCH effects). The FX observations of USD/DEM and USD/JPY are no exception. The univariate normality of the deseasonalised returns can be formally tested using the Jarque–Bera test [47]. The corresponding normal QQ–plots were ploted in Figures 2.7 and 2.8. The Jarque–Bera test is based on the fact that a normally distributed random variable has skewness zero and kurtosis equal to three. The test statistic is

$$T = \frac{n}{6}\left(\widehat{\mu_3/\sigma^3}^2 + \frac{\left(\widehat{\mu_4/\sigma^4} - 3\right)^2}{4}\right),$$

where $\widehat{\mu_3/\sigma^3}$ is the sample skewness, $\widehat{\mu_4/\sigma^4}$ is the sample kurtosis and n is the sample size. As usual μ_i denotes the ith central moment and σ the standard deviation. Under the null hypothesis that the data are normally distributed, T has asymptotically a chi–square distribution with two degrees of freedom. We ran the Jarque–Bera test and normality is rejected for all considered time series of USD/DEM and USD/JPY deseasonalised returns; see Table 3.1.

	USD/DEM		USD/JPY	
Frequency	Test statistic	P–value	Test statistic	P–value
1 hour	114 785.4	0.0	174 000.0	0.0
2 hours	78 207.3	0.0	95 757.0	0.0
4 hours	24 619.7	0.0	24 551.5	0.0
8 hours	6 246.1	0.0	10 276.3	0.0
12 hours	1 668.0	0.0	3 594.3	0.0
1 day	391.3	0.0	1 094.2	0.0

Table 3.1: *Jarque–Bera test statistic values and the p–values for the USD/DEM and USD/JPY deseasonalised returns at six frequencies.*

In order to test for conditional heteroscedasticity we recall first the definition of an ARCH process. Consider the sequence of random variables

$(X_t)_{t\in\mathbb{Z}}$ described by

$$\begin{aligned} X_t &= \mu + \epsilon_t \\ \epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2, \end{aligned} \tag{3.1}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid random variables with zero mean and unit variance. Moreover, $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, 2, \ldots, p$ and Z_t is independent of $(X_s)_{s\leq t}$ for all t. The set of equations (3.1) define what Engle [31] introduced as an autorregressive conditional heteroscedastic process of order p and it is referred to as the ARCH(p) process, where $p \in \mathbb{N}$. Given that ϵ_t has mean zero for all t, provided that X_t has second moment,

$$Var(\epsilon_t) = E(\epsilon_t^2) = \sigma_t^2,$$

where *Var* and E denote respectively the variance and the expected value both conditional on the information at time $t-1$. Then, the last equation of (3.1) can be rewritten as

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 + U_t \tag{3.2}$$

where $U_t = \epsilon_t^2 - E\left(\epsilon_t^2\right)$ is a white noise process with mean zero. Equation (3.2) shows that the squared residuals of an ARCH(p) process follow an autoregressive process of order p. In Engle [31], the author proposes a test for ARCH effects based on equation (3.2). In order to test for ARCH effects, one tests the null hypothesis: $\alpha_1 = \alpha_2 \ldots = \alpha_p = 0$ with the test statistic $LM = nR^2$. Under the null hypothesis LM has an approximate chi–square distribution with p degrees of freedom. R^2 is the coefficient of determination of the regression of ϵ_t^2 on an intercept and p lagged values of ϵ_t^2. We obtain R^2 from (3.2) after fitting the model corresponding to (3.1).

At all the frequencies the test for absence of ARCH is rejected for the USD/DEM and USD/JPY deseasonalised returns; see Table 3.2 where $p = 12$.

In our discrete–time setting, we model stochastic volatility effects by ARMA–GARCH models; see Shephard [77] for an overview on volatility

	USD/DEM		USD/JPY	
Frequency	Test statistic	P–value	Test statistic	P–value
1 hour	2 893.13	0.0	3 083.30	0.0
2 hours	857.24	0.0	1 178.70	0.0
4 hours	578.51	0.0	614.08	0.0
8 hours	190.61	0.0	336.04	0.0
12 hours	117.80	0.0	160.43	0.0
1 day	35.25	0.0	75.99	0.0

Table 3.2: *ARCH effects test statistic values and the p–values for the USD/DEM and USD/JPY deseasonalised returns at six frequencies.*

models. Formally, consider the sequence of iid random variables with zero mean and unit variance $(Z_t)_{t\in\mathbb{Z}}$. The process $(X_t)_{t\in\mathbb{Z}}$ is an ARMA(p_1, q_1)–GARCH(p_2, q_2) if it satisfies the equations

$$\begin{aligned} X_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu + \sum_{i=1}^{p_1} \phi_i \left(X_{t-i} - \mu\right) + \sum_{j=1}^{q_1} \theta_j \epsilon_{t-j} \\ \epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p_2} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{3.3}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, 2, \ldots, p_2$, $\beta_j \geq 0$ for $j = 1, 2, \ldots, q_2$ and Z_t is independent of $(X_s)_{s\leq t}$. The polynomials $\phi(z) = 1 - \phi_1 z - \ldots - \phi_{p_1} z^{p_1}$ and $\theta(z) = 1 - \theta_1 z - \ldots - \theta_{p_1} z^{q_1}$ have no common roots and no roots inside the unit circle. See [9] and [26] for more details.

We fitted univariate ARMA–GARCH models by maximum likelihood to each of the marginal series assuming that the innovations Z_t come from a t–distribution with ν degrees of freedom. Table 3.3 gives the order of the models fitted and the estimates of ν for the t–innovations. Note that, the t–distributions fitted at the one, two and four hours frequencies have infinite kurtosis because $\hat{\nu} < 4$ and so the fourth moment does not exist. The number of parameters estimated within the twelve models sums up to 117. We do not list here these parameter estimates and standard errors because our main goal is to stress the dependence structure analysis.

USD/DEM

Frequency	p_1	q_1	p_2	q_2	$\hat{\nu}$ (s.ê.)
1 hour	–	–	1	1	3.693 (0.054)
2 hours	2	2	2	1	3.708 (0.044)
4 hours	–	5	1	1	3.975 (0.105)
8 hours	2	4	1	1	4.679 (0.234)
12 hours	1	–	1	1	5.385 (0.326)
1 day	1	–	1	1	5.797 (0.556)

USD/JPY

Frequency	p_1	q_1	p_2	q_2	$\hat{\nu}$ (s.ê.)
1 hour	–	–	1	1	3.654 (0.052)
2 hours	1	–	2	1	3.759 (0.077)
4 hours	4	4	2*	1	3.819 (0.109)
8 hours	2	2	1*	1	4.357 (0.195)
12 hours	1	–	1*	1	4.574 (0.251)
1 day	10	–	1*	1	4.889 (0.412)

Table 3.3: *Order of the ARMA(p_1, q_1)–GARCH(p_2, q_2) models fitted to the USD/DEM and USD/JPY returns at the several frequencies. Degrees of freedom estimated for the marginal conditional distribution t of the innovations and corresponding standard errors are also given.*

In each univariate model we included a leverage effect parameter γ in the GARCH dynamics; see for example Bollerslev et al. [4] and references therein. The introduction of γ attempts to take into account an asymmetric contribution that the innovations may have on the volatility. This improvement is also possible in the model specified in (3.3); see Ding et al. [20] and Zivot and Wang [82] where the GARCH component of model (3.3) is treated as a special case of a power GARCH model. In this case, the last equation in (3.3) becomes

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i})^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2.$$

We use the usual t–statistic $\hat{\alpha}/\hat{\sigma}_{\hat{\alpha}}$ to test whether the general model parameter α is zero. For the USD/DEM returns we can not reject the null hypothesis of $\gamma_k = 0$ for the estimated γ parameter and this at all frequencies. In the case of USD/JPY the situation is almost the complete reverse. In this case, we reject the null hypothesis for all frequencies lower

than (and including) four hours. In Table 3.3 we put an superscript * in the ARCH order of such frequencies. Rejecting the null hypothesis for the USD/JPY model parameter $\gamma_k = 0$, and using that the estimated values $\hat{\gamma}$ are negative, we have that negative shocks (bad news) have a larger impact on volatility than positive shocks (good news).

From the fitted ARMA–GARCH model parameters we recover the residuals or filtered returns $\hat{z}_t$ for each univariate time series. The residuals are computed from the sample $(x_1, x_2, \ldots, x_n)$ and from the fitted parameters as

$$\hat{z}_t = (x_t - \hat{\mu}_t)/\hat{\sigma}_t \tag{3.4}$$

for $t = 1, 2, \ldots, n$. In Figures 3.1 and 3.2 we plot the sample autocorrelograms and cross–correlograms for the absolute values of the bivariate filtered one and eight hours FX returns respectively; namely absolute values of the USD/DEM and USD/JPY residual vectors $\hat{z}_t$ resulting from the above marginal ARMA–GARCH fitting. Especially in Figure 3.2 there is no evidence against serial independence of the absolute residual values. Comparing Figure 3.1 with Figure 2.6 we can also find less serial correlation in the filtered returns. Only the contemporaneous cross–dependence definitely remains (see lag zero in the cross–correlograms of Figures 3.1 and 3.2) and that is exactly where our interest lies.

Though there exist several multivariate GARCH models in the literature, like CCC–GARCH, DVEC, matrix–diagonal GARCH, BEKK, principal components GARCH, in our first analysis we did not want to bias our investigation of the dependence structure by imposing a specific analytic model on it. In Chapter 4, we will reanalyse the data using a matrix–diagonal GARCH model.

We proceed in this chapter by performing the copula analysis of the bivariate residuals or filtered returns $\hat{\mathbf{z}}_t$. We basically repeat, adding some more copula models, the analysis done in Chapter 2 on the bivariate deseasonalised USD/DEM and USD/JPY returns $\mathbf{x}_t$, but now on the residuals $\hat{\mathbf{z}}_t$.

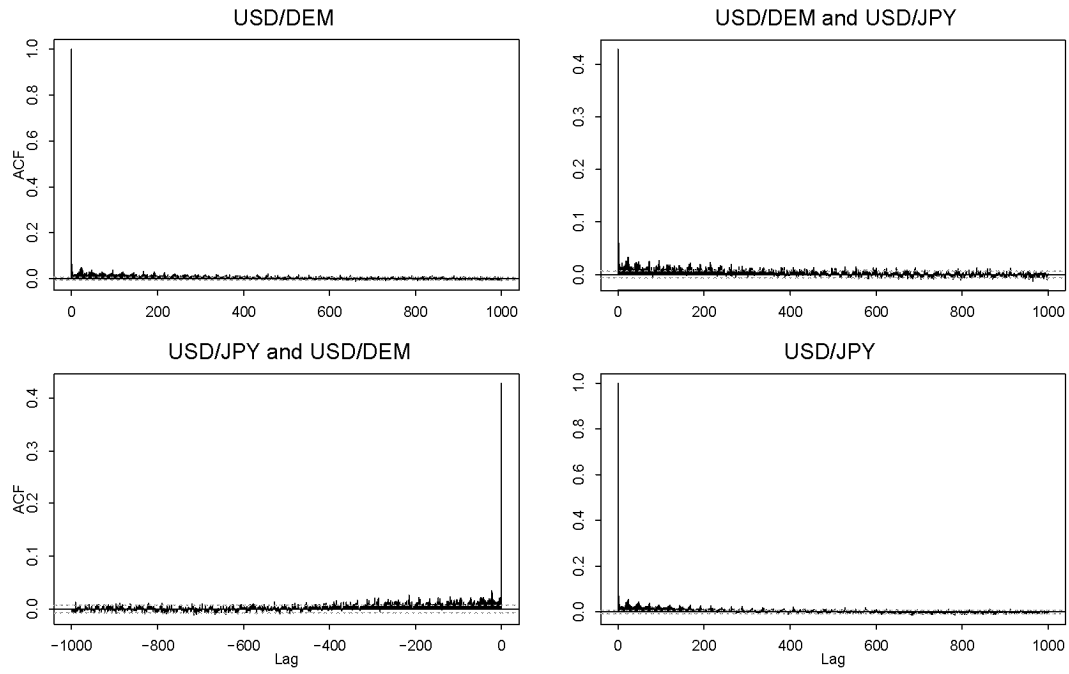

Figure 3.1: *Sample autocorrelograms for the absolute values of the one hour USD/DEM and USD/JPY residuals, respectively top left and bottom right, and cross–correlograms for USD/DEM on past USD/JPY (top right) and USD/JPY on past USD/DEM (bottom left).*

3.2 Copulae for USD/DEM and USD/JPY residuals

Figure 3.3 shows the scatter–plots of the USD/DEM and USD/JPY residuals $\hat{\mathbf{z}}_t$ obtained through the fitting of Section 3.1. Suppose that, for a given time frequency, the USD/DEM residuals are represented by the random variable Z_1 and the USD/JPY by the random variable Z_2. Assume that (Z_1, Z_2) has multivariate distribution function F and continuous univariate marginal distribution functions F_1 and F_2. In order to investigate the residual dependence, we fit copula–based models of type (1.1).

For a fixed frequency, given a sample of filtered returns

$$(\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2, \ldots, \hat{\mathbf{z}}_n),$$

the dependence parameter $\boldsymbol{\theta}$ of the copula C is estimated by the pseudo

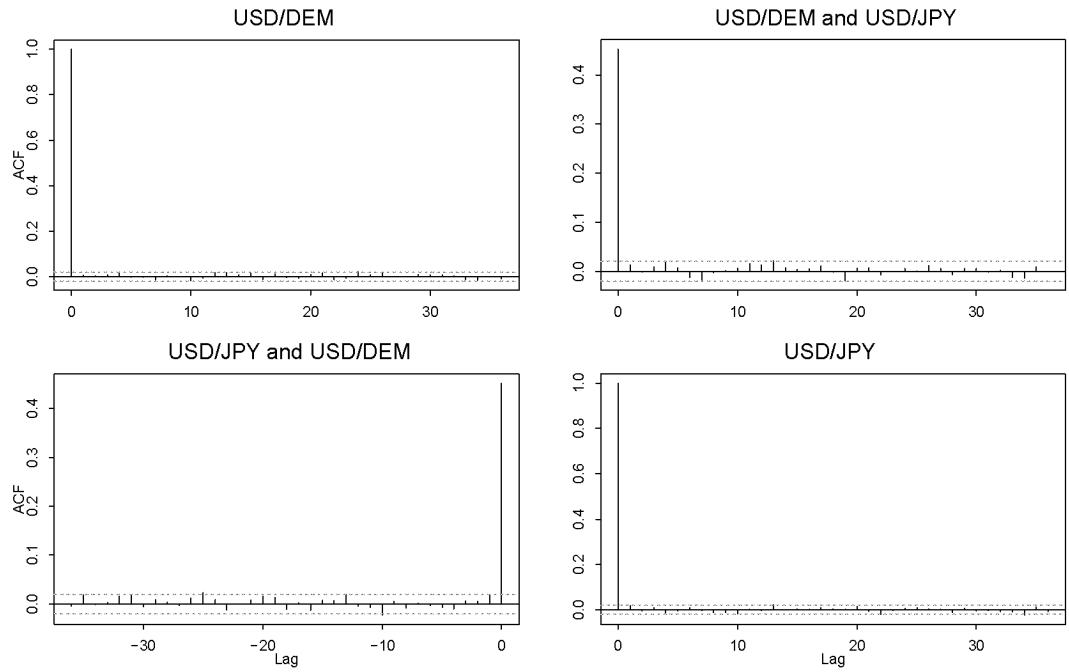

Figure 3.2: *Sample autocorrelograms for the absolute values of the eight hours USD/DEM and USD/JPY residuals, respectively top left and bottom right, and cross–correlograms for USD/DEM on past USD/JPY (top right) and USD/JPY on past USD/DEM (bottom left).*

log–likelihood estimator from Section 1.4 by maximisation of (1.18) which now reads

$$L(\boldsymbol{\theta}; \hat{\mathbf{z}}) = \sum_{i=1}^{n} \log c(F_{1n}(\hat{z}_{1i}), F_{2n}(\hat{z}_{2i}); \boldsymbol{\theta}).$$

We display in Figure 3.4 the contour–plots of the bivariate filtered returns mapped into the unit square by its marginal empirical distributions. We should compare these graphs with the contour–plots of the deseasonalised returns in Figure 2.11. From the contour–plots we may conclude that there are no evident changes in the contour shapes between the returns and the filtered returns. This is in contrast with the differences observed in the serial correlations between the returns and the residuals comparing Figure 2.6 with Figures 3.1 and 3.2.

The copula families fitted to the USD/DEM and USD/JPY spot rate residuals are: t, Frank, Plackett, Gaussian, Gumbel, Clayton and the mixtures Gumbel with survival Gumbel, Clayton with survival Clayton, Gum-

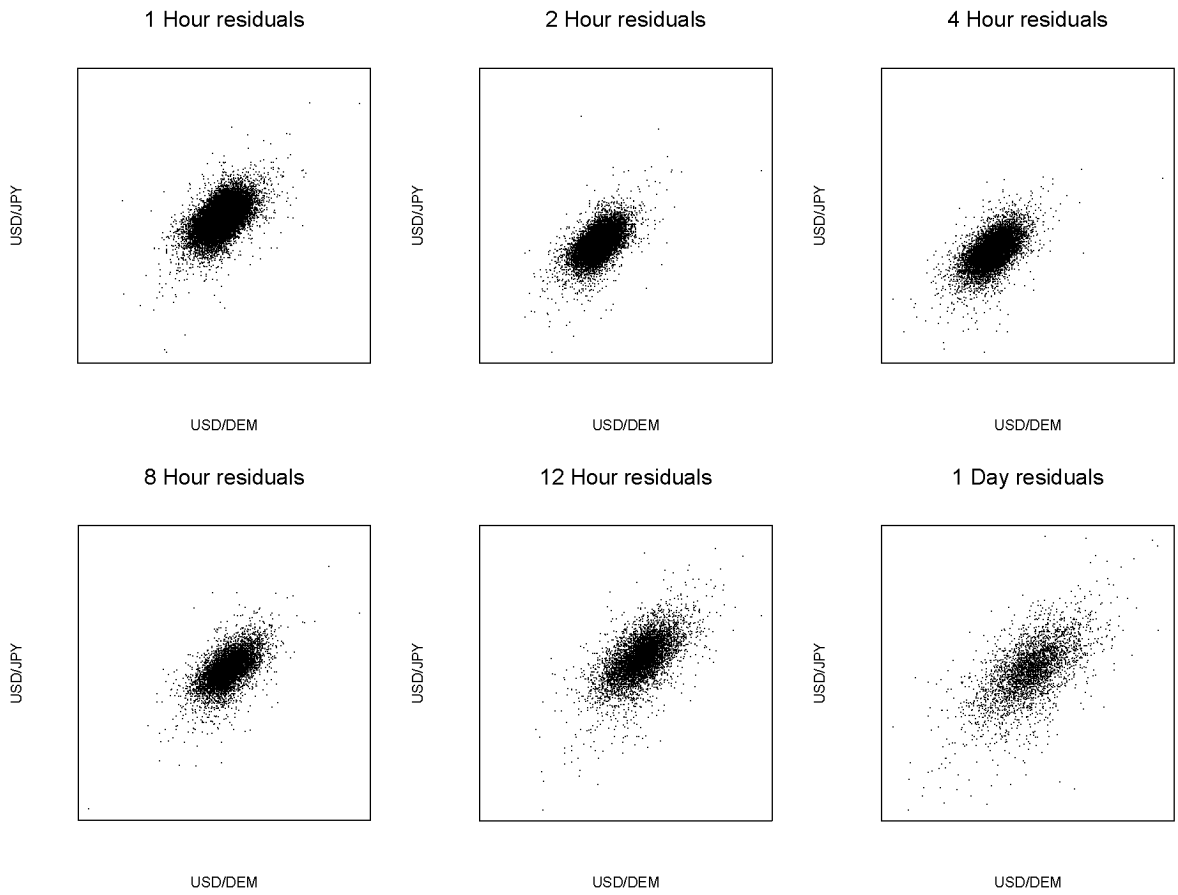

Figure 3.3: *FX spot rates for USD/DEM and USD/JPY. The figure displays the scatter–plots of the filtered returns for the several time frequencies.*

bel with Clayton and survival Gumbel with survival Clayton. This choice of time–invariant copula models is partly based on the previous analysis but also on tractability and flexibility for investigating tail–dependence.

We fitted all the listed copula models to the USD/DEM and USD/JPY residuals to obtain the dependence parameter estimates $\hat{\boldsymbol{\theta}}$ for the several frequencies. As before, the models are ranked by their Akaike information value. Secondly, the goodness of fit test from Section 1.6 is performed to the best ranked models for each time frequency. Parameter estimates and the approximated standard errors ($s\hat{e}.$) for all fitted models are listed in Tables 3.4 and 3.5. For the t–copula the parameters θ_1 and θ_2 in Table 3.5 represent respectively the degrees of freedom and the correlation.

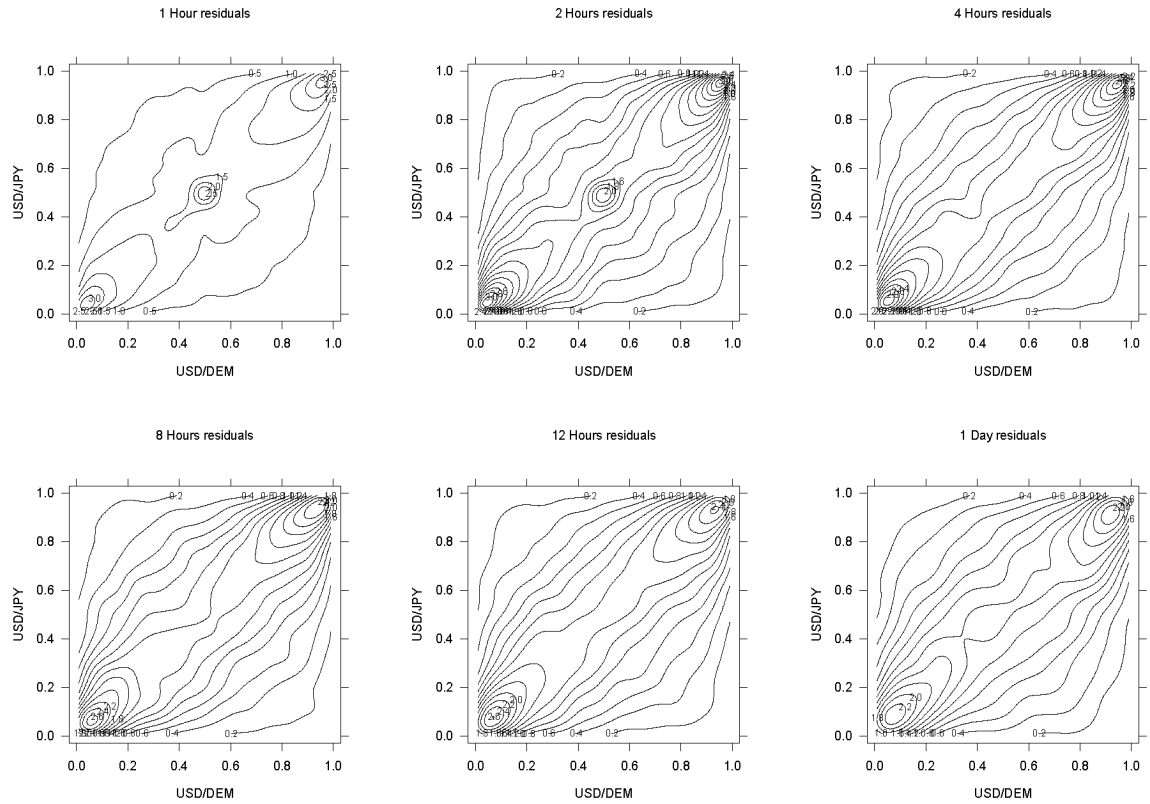

Figure 3.4: *Contour–plots of the bivariate filtered returns of USD/DEM and USD/JPY for different time frequencies mapped into the unit square by its marginal empirical distributions.*

From the models fitted to the residuals, the one which has the best AIC is the t–copula for almost all the frequencies. The exception are the daily observations where the mixture of roughly 0.5 of Gumbel with 0.5 survival Gumbel performs slightly better than the t–model. To enable an easy comparison among the AIC values obtained for each frequency we again plotted in Figure 3.5 the relative differences between the AIC for the t–model and the AIC for all the other models. The results for the Clayton and survival Clayton are not included in the plot because they are significantly worse for all time frequencies.

According to the AIC criterion, the mixture models and the t–model perform better than the one–parameter models. Both copulae with very asymmetric tails give poor fits. We should remark that the Plackett copula is the best of the one parameter models except for daily residuals

where the Gaussian is better. The Gaussian and the t–copula have very different AIC values. Note that the t–copula having degrees of freedom estimated between $\hat{\nu} = 4.7$ and $\hat{\nu} = 6$ is still far from its Gaussian limit. Nevertheless, the t copula approaches the Gaussian for a decreasing time frequency (central limit effect).

From this analysis we can conclude that the filtered residuals on USD/DEM and USD/JPY spot rates can be modelled well by the t–model or by a mixture between the Gumbel and survival Gumbel copulae. These are always the two best models.

Comparing these results with those obtained previously using an unconditional, static iid model in Chapter 2 the conclusions are similar. There the t–model was always the best for the several frequencies of returns; the Gumbel mixture model was however not included.

Finally for this section we give in Table 3.6 the p–values for the two best copula–based models for each time frequency. The reported p–values are computed using the probability integral test as discussed in Section 1.6.

In the last two sections we did a first dynamic modelling of the dependence structure of the USD/DEM and USD/JPY returns. We can summarise the modelling procedure as follows: we fit univariate ARMA–GARCH models corresponding to (3.3) to each marginal vector of the bivariate samples $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ of deseasonalised returns at all frequencies. As an outcome, we compute the filtered returns $(\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2, \ldots, \hat{\mathbf{z}}_n)$ through (3.4). Then we model the dependence structure of these residuals with a copula–based model (1.1).

Frequency	Copula model	$\hat{\theta}$ $(\hat{s.e.})$	AIC
1 hour	Clayton	0.859 (0.006)	-23401.101
	Frank	3.979 (0.024)	-27032.306
	Gaussian	0.550 (0.002)	-28267.108
	Gumbel	1.562 (0.004)	-28146.727
	Plackett	6.503 (0.061)	-29324.002
2 hours	Clayton	0.913 (0.009)	-12730.906
	Frank	4.200 (0.035)	-14806.051
	Gaussian	0.571 (0.002)	-15483.028
	Gumbel	1.605 (0.006)	-15506.480
	Plackett	7.038 (0.093)	-16020.855
4 hours	Clayton	0.944 (0.013)	-6652.147
	Frank	4.341 (0.050)	-7821.259
	Gaussian	0.584 (0.004)	-8176.122
	Gumbel	1.634 (0.009)	-8251.189
	Plackett	7.361 (0.137)	-8446.380
8 hours	Clayton	0.984 (0.019)	-3536.107
	Frank	4.563 (0.072)	-4260.632
	Gaussian	0.603 (0.005)	-4413.271
	Gumbel	1.669 (0.013)	-4412.292
	Plackett	7.752 (0.201)	-4533.552
12 hours	Clayton	1.025 (0.024)	-2487.922
	Frank	4.659 (0.088)	-2941.280
	Gaussian	0.615 (0.006)	-3092.874
	Gumbel	1.681 (0.016)	-3007.219
	Plackett	7.949 (0.252)	-3113.152
1 day	Clayton	1.034 (0.035)	-1252.289
	Frank	4.599 (0.124)	-1446.464
	Gaussian	0.617 (0.009)	-1552.695
	Gumbel	1.679 (0.023)	-1500.065
	Plackett	7.772 (0.350)	-1526.993

Table 3.4: *Residuals on USD/DEM and USD/JPY log–returns. Estimates and standard errors of dependence parameters in Clayton, Frank, Gaussian, Gumbel and Plackett models. For each model fitted we provide the AIC value. The reading of this table must be complemented with Table 3.5.*

Freq.	Copula model	$\hat{\theta}_1$ $(\hat{s.e.})$	$\hat{\theta}_2$ $(\hat{s.e.})$	$\hat{\theta}_3$ $(\hat{s.e.})$	AIC
1 hour	Cl & s. Cl	1.125 (0.025)	1.171 (0.027)	0.516 (0.007)	-29924.80
	Cl & Gumbel	1.568 (0.014)	1.363 (0.066)	0.659 (0.009)	-30642.91
	s.Cl & s.Gum	1.552 (0.010)	1.510 (0.065)	0.701 (0.008)	-30665.31
	Gum & s.Gum	2.038 (0.030)	1.405 (0.009)	0.421 (0.010)	-31061.35
	t	4.935 (0.108)	0.558 (0.002)	–	-31517.70
2 hours	Cl & s. Cl	1.164 (0.033)	1.316 (0.041)	0.517 (0.010)	-16430.36
	Cl & Gumbel	1.674 (0.034)	1.233 (0.114)	0.650 (0.014)	-16803.66
	s.Cl & s.Gum	1.576 (0.013)	1.723 (0.086)	0.695 (0.011)	-16801.39
	Gum & s.Gum	2.109 (0.039)	1.420 (0.013)	0.441 (0.013)	-17015.86
	t	4.822 (0.147)	0.580 (0.003)	–	-17192.73
4 hours	Cl & s. Cl	1.238 (0.048)	1.325 (0.051)	0.499 (0.014)	-8653.704
	Cl & Gumbel	1.682 (0.032)	1.359 (0.128)	0.674 (0.017)	-8863.199
	s.Cl & s.Gum	1.640 (0.024)	1.535 (0.115)	0.669 (0.016)	-8847.032
	Gum & s.Gum	1.501 (0.028)	1.991 (0.064)	0.545 (0.020)	-8932.445
	t	4.748 (0.201)	0.593 (0.004)	–	-9088.884
8 hours	Cl & s. Cl	1.265 (0.060)	1.472 (0.071)	0.502 (0.018)	-4607.028
	Cl & Gumbel	1.771 (0.053)	1.265 (0.170)	0.667 (0.024)	-4722.141
	s.Cl & s.Gum	1.663 (0.028)	1.710 (0.140)	0.668 (0.021)	-4713.239
	Gum & s.Gum	1.991 (0.072)	1.534 (0.040)	0.496 (0.027)	-4764.398
	t	5.323 (0.343)	0.612 (0.006)	–	-4818.328
12 hours	Cl & s. Cl	1.492 (0.095)	1.286 (0.084)	0.503 (0.024)	-3157.357
	Cl & Gumbel	1.653 (0.031)	1.893 (0.184)	0.679 (0.025)	-3242.862
	s.Cl & s.Gum	1.787 (0.060)	1.307 (0.206)	0.673 (0.030)	-3248.558
	Gum & s.Gum	1.556 (0.046)	2.018 (0.088)	0.511 (0.033)	-3281.252
	t	5.837 (0.505)	0.621 (0.007)	–	-3304.250
1 day	Cl & s. Cl	1.548 (0.120)	1.280 (0.099)	0.494 (0.032)	-1599.798
	Cl & Gumbel	1.665 (0.045)	1.844 (0.249)	0.671 (0.037)	-1629.394
	s.Cl & s.Gum	1.816 (0.071)	1.234 (0.195)	0.656 (0.039)	-1632.435
	Gum & s.Gum	1.588 (0.072)	1.952 (0.117)	0.501 (0.048)	-1642.460
	t	6.012 (0.786)	0.620 (0.010)	–	-1640.061

Table 3.5: *Residuals on USD/DEM and USD/JPY log–returns. Estimates and standard errors of parameters for the t–model and for the four mixture models considered. In case of the mixture models, θ_1 and θ_2 are the dependence parameters respectively for the first and second terms of the mixture. θ_3 is the mixture parameter which gives the proportion of the first term. For the t–model, θ_1 are the degrees of freedom and θ_2 is the correlation. For each model fitted we provide the AIC value. The reading of this table must be complemented with Table 3.4.*

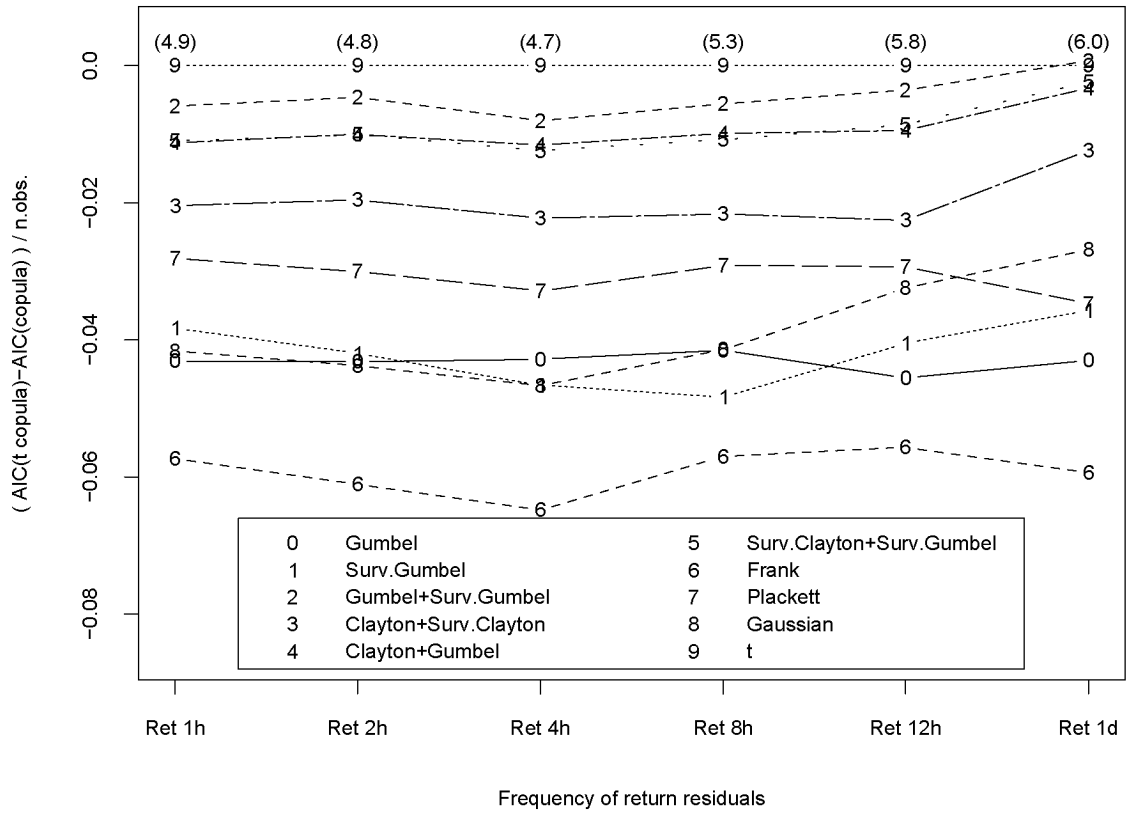

Figure 3.5: *Plot of the AIC values relative to the t–copula and to the sample size for each model and time frequency.*

3.3 Tail–dependence coefficient

As in Section 2.2 of Chapter 2, the fitted models are now used to estimate the tail–dependence coefficient, λ defined in Section 1.2; in this case for the filtered data. To alow for some comparison we compute λ for the two best models, in terms of AIC. Recall that these models are the t–copula and the mixture of Gumbel with survival Gumbel copulae. The tail–dependence coefficient for the t–copula was given in Example 1.15 and and for the mixture model in (1.35).

Table 3.7 has the estimated values for the tail–dependence coefficients.

Frequency	sample size	t–model	Gumbel mixture
1 hour	78,239	0	0
2 hours	39,119	0	0
4 hours	19,559	0.0348	0.0006
8 hours	9,779	0.3808	0.1079
12 hours	6,519	0.2471	0.1949
1 day	3,259	0.7211	0.6775

Table 3.6: *P–values for a goodness–of–fit test of the fitted t and Gumbel mixture models to the residual returns on USD/DEM and USD/JPY spot rates.*

Here again we could use jackknife or bootstrap with Monte Carlo simulation to estimate confidence intervals for the λ estimators, like explained in Section 1.8; we did not include these results.

	t copula	Gumbel mixture	
Frequency	λ	λ_L	λ_U
1 hour	0.242	0.209	0.250
2 hours	0.261	0.207	0.269
4 hours	0.273	0.265	0.225
8 hours	0.261	0.216	0.289
12 hours	0.247	0.288	0.224
1 day	0.240	0.286	0.226

Table 3.7: *Lower and upper tail–dependence coefficients for the residual returns on USD/DEM and USD/JPY spot rates given by the fitted t and Gumbel mixture models.*

The results in Table 3.7 are good in the sense that both models give very similar tail–dependence coefficient estimates. From the values obtained we can say that tail–dependence is still present in the residuals. Comparing the estimated tail–dependence coefficient of the t models fitted to the deseasonalised returns (see Table 2.2) and to the filtered returns (see Table 3.7) we find no significant differences.

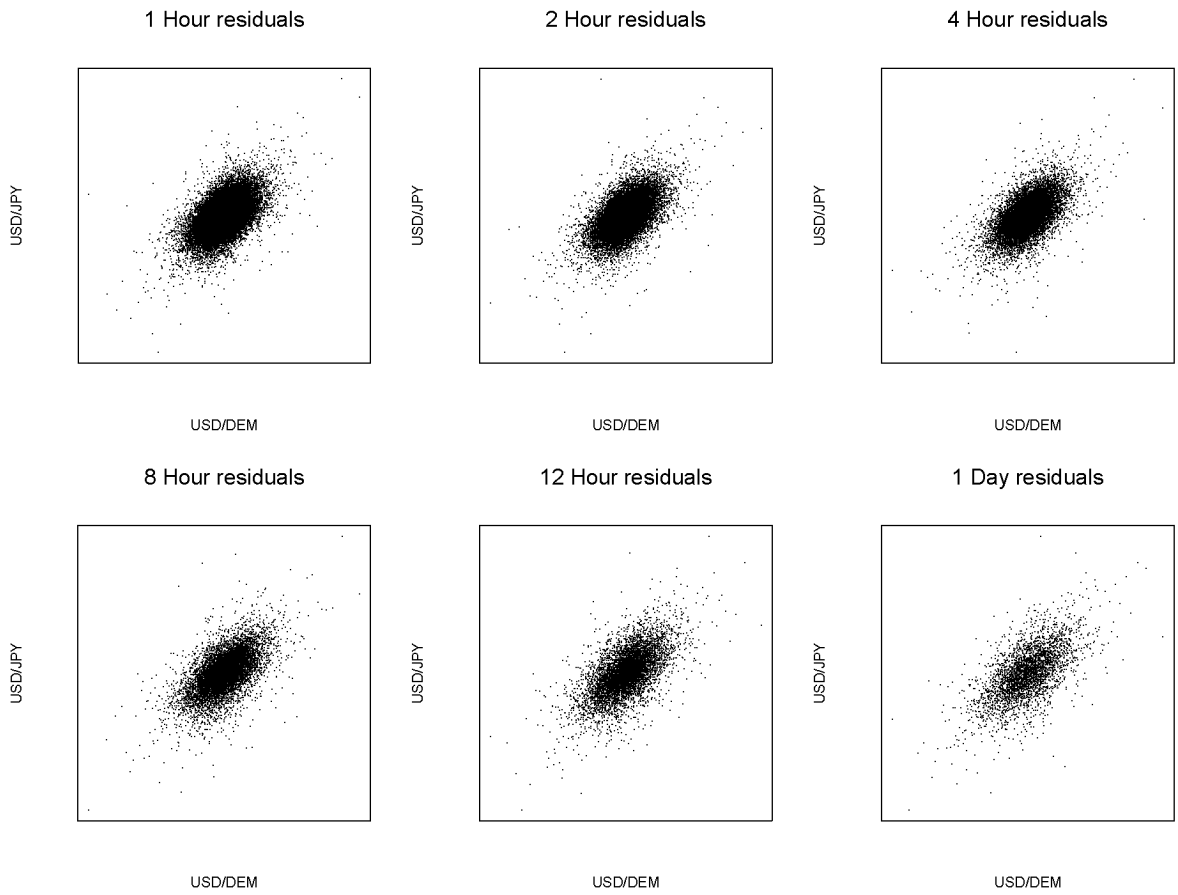

Figure 3.6: *FX spot rates for USD/DEM and USD/JPY. The figure displays the scatter–plots of the filtered returns for the several time frequencies and with margins transformed into standard t–distributed observations with the degrees of freedom estimated for the corresponding t–copula.*

3.4 Testing for ellipticity

Whereas in Chapter 2 we tested on the ellipticity of the deseasonalised log–return data itself, now we will perform such a test on the residuals. We use again the test discussed in Section 2.3.1. First we test ellipticity for the original residuals $\hat{\mathbf{z}}_t$ and secondly for the residuals where we transform the margins into standard univariate t–distributed observations using the degrees of freedom estimated for the corresponding t–copula, given in Table 3.5,

$$\left(t_{\hat{\nu}}^{-1}\left(F_{1n}\left(z_{1i}\right)\right), t_{\hat{\nu}}^{-1}\left(F_{2n}\left(z_{2i}\right)\right)\right)$$

for $i = 1, 2, \ldots, n$ and for the six frequencies, like in (2.10).

Table 3.8 has the p–values obtained for the ellipticity test using the two kinds of margins. The second column has the results of the test on the original residuals plotted in Figure 3.3 and the third column has the results of testing on the marginal t–distributed residuals, displayed in Figure 3.6.

For the original margins, the ellipticity hypothesis is rejected from one hour up to eight hours periods and not rejected for twelve hours and one day frequencies. After transforming the margins, ellipticity is rejected for one hour and two hours frequencies and not rejected for all the remaining time horizons. Compared with the results obtained in Section 2.3.2 for the returns without filtering we can say that we "gained" one frequency. Before, the ellipticity was also rejected for the twelve hours period with original margins and for the four hours frequency with transformed margins. This may be due to the gain of information resulting from passing from clearly non–independent samples of returns to much less dependent samples of filtered returns. Also here we stress the fact that our test results are based on varying sample sizes.

Frequency	Original margins	t margins
1 hour	0	0
2 hours	0	0
4 hours	0	0.348
8 hours	0.001	0.069
12 hours	0.145	0.501
1 day	0.389	0.451

Table 3.8: *P–values for the ellipticity test for the filtered returns on USD/DEM and USD/JPY spot rates with the original and with the t–transformed margins.*

We further estimated, on the residuals, the spectral measure and the limit behaviour of high (low) level bivariate excesses. The results obtained are very similar to those reported in Chapter 2 for the return data itself. We hence will not report in detail these results here but rather turn to a more careful bivariate dynamic modelling in Chapter 4.

Comments

Starting from the deseasonalised returns on USD/DEM and USD/JPY from Section 2.1 we analysed, at all frequencies, the dependence structure of the data considering the univariate serial dependence. While in Section 2.2 we ignored this fact, in this chapter we modelled the contemporaneous dependence after filtering the returns using univariate ARMA–GARCH models. The bivariate residuals of returns so obtained showed no strong evidence against the iid property but were still contemporaneously dependent. Copula–based models were fitted to the full data sets of residuals at each frequency. The various results obtained were fairly similar to those obtained in Chapter 2.

Chapter 4

Time–varying copula models

On dealing with multivariate risks it is well known that the homogeneity of the marginal behaviour of each risk does not imply a global homogeneous behaviour. For example, there is considerable interest in the dynamic behaviour of correlation between different risks as a function of time; see for instance Boyer et al. [7], Longin and Solnik [57] and Loretan and Phillips [58]. Because of the fundamental importance of the notion of linear correlation in finance and insurance, such dynamics may have a non–trivial impact on the pricing and hedging of underlying instruments, or of the risk measurement thereof. As a consequence, a more systematic modelling for the dynamic behaviour of the dependence structure underlying multivariate risks is called for. This chapter and the next one attempt to answer this question using two different approaches.

As we saw in the previous chapter, for the FX spot rates on USD/DEM and USD/JPY the contemporaneous conditional dependence structure between the two residual series is well described by a t–copula or by a mixture of a Gumbel with a survival Gumbel copula. In Chapter 3 the uni-

variate serial dependence was modeled by ARMA–GARCH models with t–innovations. These two aspects of dependence were modelled independently in two steps. First the univariate time series models were fitted and then time–invariant copula–based models were used to model the cross–dependence structure of the residuals resulting from the marginal modelling. In this chapter, we model the dynamics of the time dependence structure as well as the dynamics of the contemporaneous dependence. For that we want to combine two univariate ARMA–GARCH models with a time–varying copula model. This is achieved using a copula–based model for the conditional bivariate innovations coupling two ARMA–GARCH processes. With such a procedure we investigate the constancy of the conditional dependence structure allowing for time–varying dependence parameters and assuming a fixed copula family. As before, we look at several time frequencies for the spot rates considered. In Section 4.1 we inspect the possible existence of stochastic behaviour in the dependence structure. For that we fit matrix–diagonal GARCH type models to the bivariate deseasonalised FX returns and look at the estimated time–varying correlation. As these estimates rule out the possibility of a constant correlation and consequently of a constant dependence structure, in Section 4.2 we propose a multivariate dynamic model where the copula is time–varying. The matrix–diagonal model imposes on the innovations a Gaussian copula or a t–copula with the same degrees of freedom for the margins. The model presented allows to use any copula to link the univariate innovations. In Section 4.3 we explain how to estimate the parameters of the time–varying copula–based model. We apply this model to the several frequencies of FX returns on USD/DEM and USD/JPY in Section 4.4.

4.1 Stochastic dependence structure

In the econometric literature, stochastic volatility is accepted as a stylised fact in most financial univariate processes. If the conditional volatility of a series is not constant there is no reason why we should expect constancy

for the correlation between different time series or more generally for the conditional dependence structure (copula).

In order to inspect for time–varying cross–correlation we can construct the so called exponentially weighted covariance estimate; see Andreou and Ghysels [2] and Foster and Nelson [34]. In the same spirit, but more sophisticated, are the matrix–diagonal GARCH models as proposed by Bollerslev et al. [5]. We choose the later models as a first approach to look at the covariance component of the conditional dependence structure. We also considered the BEKK models from Engle and Kroner [32] but the increase in the number of parameters did not lead to a worthwhile fitting improvement.

In a way similar to a one–dimensional AR–GARCH process (3.3), the d–dimensional matrix–diagonal GARCH(p_2, q) process $(\mathbf{X}_t)_{t\in\mathbb{Z}}$, with an AR$(p_1)$ component, is defined as

$$\begin{aligned}
\mathbf{X}_t &= \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t \\
\boldsymbol{\mu}_t &= \boldsymbol{\mu} + \sum_{i=1}^{p_1} M_i \left(\mathbf{X}_{t-i} - \boldsymbol{\mu}\right) \\
\boldsymbol{\epsilon}_t &= \Sigma_t^{1/2} \mathbf{Z}_t \\
\Sigma_t &= A_0 A_0^t + \sum_{i=1}^{p_2} (A_i A_i^t) \otimes (\boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}_{t-i}^t) + \sum_{j=1}^{q} (B_j B_j^t) \otimes \Sigma_{t-j}
\end{aligned} \tag{4.1}$$

where A_i for $i = 0, 1, \ldots, p_2$ and B_j with $j = 1, 2, \ldots, q$ are lower triangular $d \times d$ matrices. Moreover, M_i is a full matrix in $\mathbb{R}^{d\times d}$ for $i = 1, 2, \ldots, n$ and p_1, p_2 and q are positive integers. The d–dimensional vector sequence $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ is assumed to be iid with zero mean vector and unit variances. The matrix Σ_t stands for the conditional covariance matrix of the vector $\boldsymbol{\epsilon}_t$ and $\Sigma_t^{1/2}$ is obtained through the Cholesky decomposition of Σ_t. In (4.1), $\otimes$ stands for the Hadamard product, the element by element multiplication. With these models we have the guarantee of obtaining positive semi–definite covariance matrix estimates. To evaluate the standardised residuals we compute $\Sigma_t^{-1/2}\boldsymbol{\epsilon}_t$.

We used maximum likelihood to fit bivariate AR–GARCH models with

matrix–diagonal multivariate specification ((4.1)) to the six time frequencies considered for the FX spot rate returns. For fitting purposes, the standardised residuals $\mathbf{Z}_t$ are assumed to come from a bivariate t–distribution. The order of the estimated models and the estimated degrees of freedom obtained for the t–innovations for each time frequency are given in Table 4.1. For each time frequency we plotted in Figures 4.1 and 4.2 the es-

Frequency	p_1	p_2	q	$\hat{\nu}(s\hat{.}e.)$
1 hour	1	1	1	4.888 (1.318)
2 hours	–	1	1	5.253 (4.358)
4 hours	1	1	1	5.351 (1.450)
8 hours	1	1	1	5.893 (0.930)
12 hours	1	1	1	5.943 (0.429)
1 day	1	1	1	5.998 (0.320)

Table 4.1: *Order of the matrix–diagonal models fitted and estimated degrees of freedom of the assumed t–innovations for the several time horizons of USD/DEM and USD/JPY spot rate returns.*

timated conditional cross–correlation. These are the off–diagonal element of $\hat{\boldsymbol{\epsilon}}_t$ for $t = 1, 2, \ldots, n$. The conditional correlation seems to fluctuate quite a lot, mostly around 0.6 with occasional drops to values that can be negative especially for the higher frequencies. There seems to be evidence for the existence of three regimes, and this consistently at all frequencies. The corresponding periods are first up to the end of 1989, then from the latter date till mid 1993, and finally from this date till the end of 1997. As we will see in the change–points analysis in Section 5.4 there is a drop in the correlation from 8 to 9 November 1989. This shift coincides with the start of the German unification while in June of 1997 the Asia crisis began. Assuming that the Berlin wall event (10:30 pm, 9 November 1989) caused the estimated shift from 8 to 9 November 1989 then with the matrix–diagonal model this change is visible much more precisely in Figures 4.1 and 4.2 at higher frequencies than at the lower frequencies.

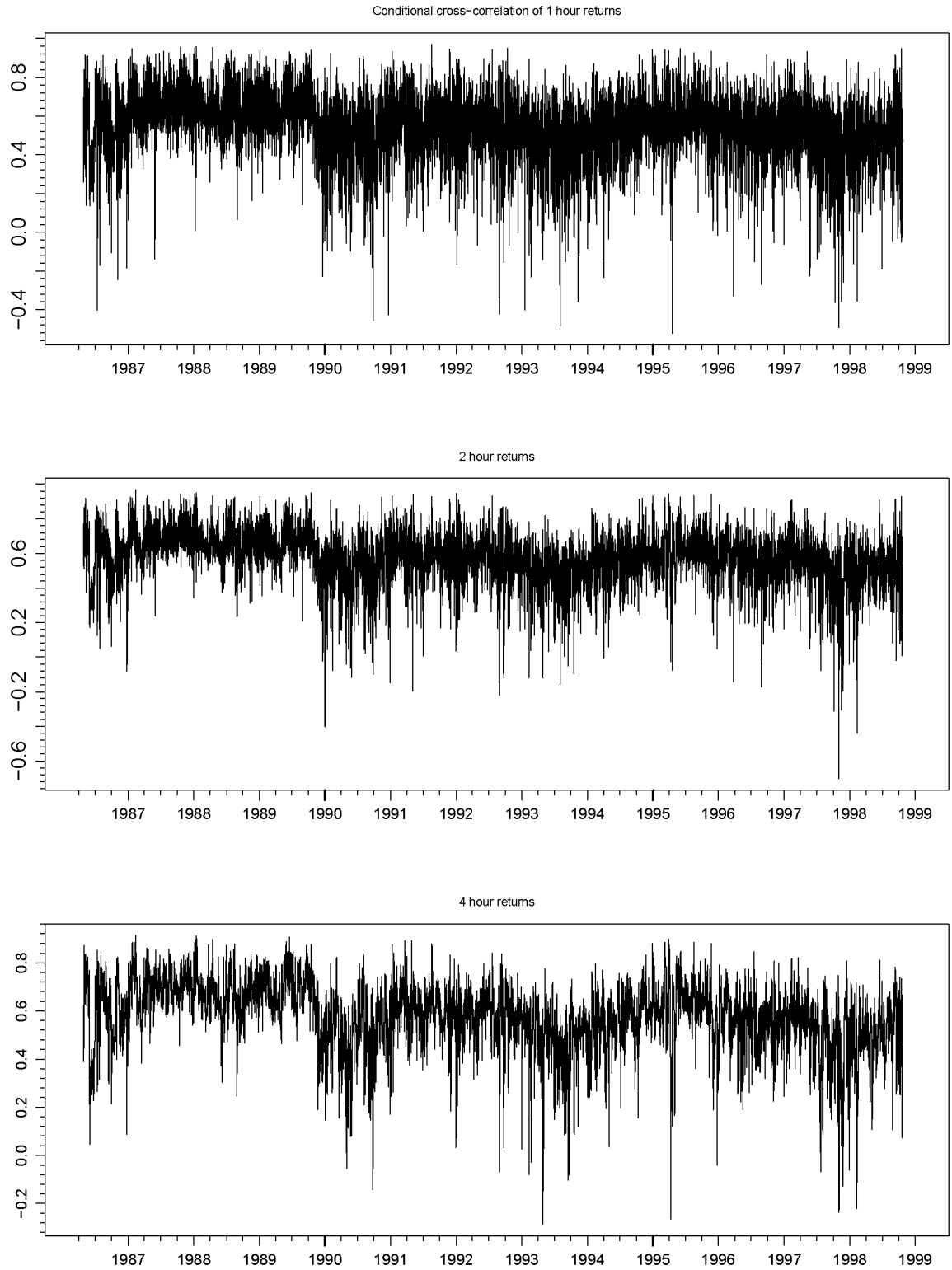

Figure 4.1: *Time–varying cross–correlations estimated by a matrix–diagonal AR–GARCH model for the returns on the FX USD/DEM and USD/JPY spot rates of one, two and four hours frequencies.*

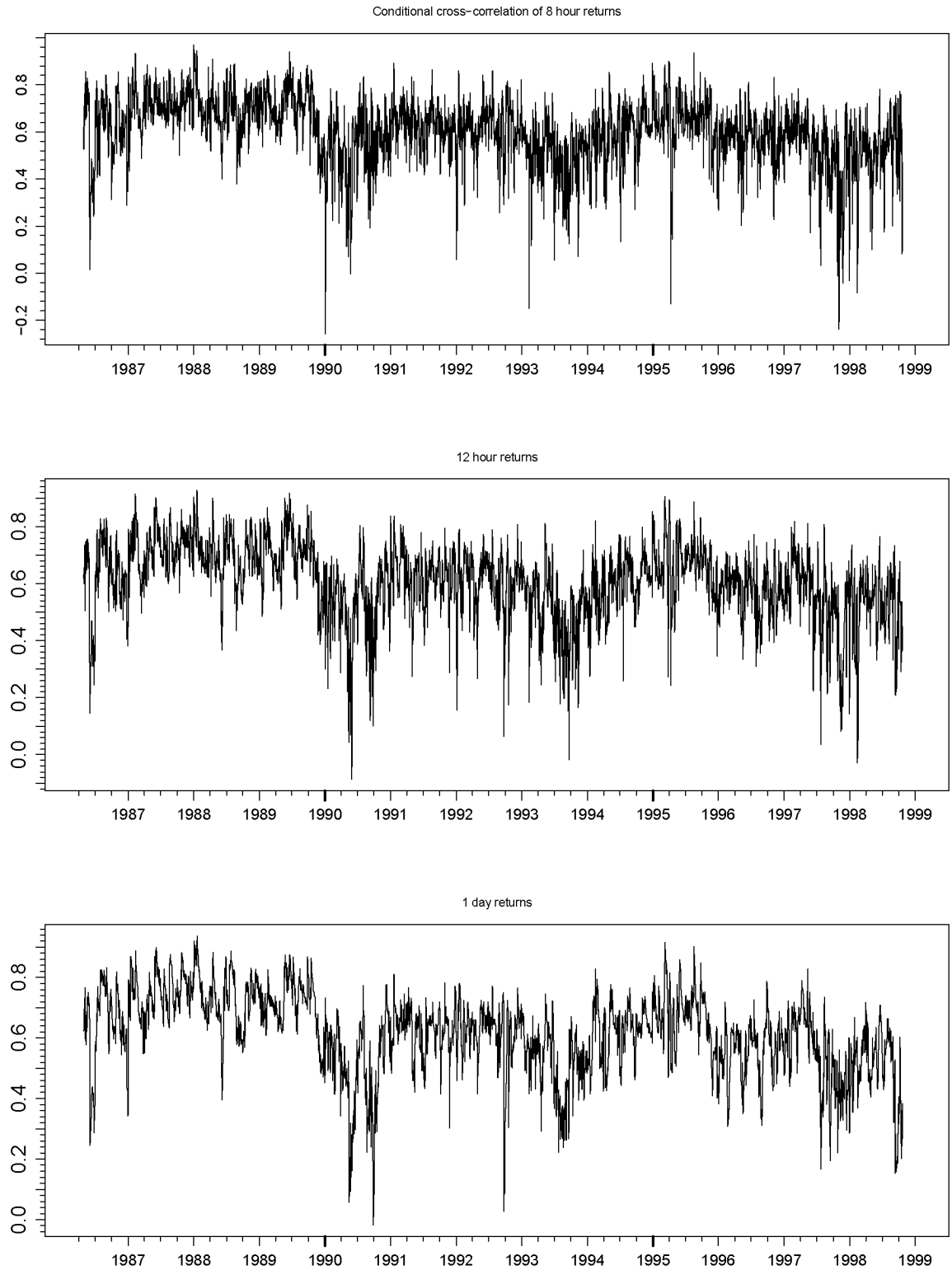

Figure 4.2: *Time–varying cross–correlations estimated by a matrix–diagonal AR–GARCH model for the returns on the FX USD/DEM and USD/JPY spot rates of eight hours, twelve hours and one day frequencies.*

4.2 The Multivariate GARCH model with time–varying copula

The estimated covariances in Section 4.1 show that we can not expect a constant type conditional dependence structure between the two return series, USD/DEM and USD/JPY. That means that we should use a model where somehow the conditional dependence variability is incorporated. Given that we want to couple two GARCH type models through a parametric copula, the simplest way is to allow for the conditional copula to stay in the same family but for the dependence parameters to be time–varying.

Let $(\mathbf{X})_{t\in\mathbb{Z}}$ be a sequence of observable d–dimensional random vectors. Consider the process description given by

$$\begin{aligned} \mathbf{X}_t &= \mathbf{c} + \boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t &= \boldsymbol{\sigma}_t \mathbf{Z}_t \\ \boldsymbol{\sigma}_t^2 &= A_0 + \sum_{i=1}^{p} A_i \otimes (\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}^t) + \sum_{j=1}^{q} B_j \otimes \boldsymbol{\sigma}_{t-j}^2 \end{aligned} \tag{4.2}$$

where A_i for $i = 0, 1, \ldots, p$ and B_j with $j = 1, 2, \ldots, q$ are diagonal $d \times d$ matrices, $\mathbf{c}$ is a vector in $\mathbb{R}^d$ and p and q are positive integers. Moreover, $(Z_{i,t})_{t\in\mathbb{Z}}$ for $i = 1, 2, \ldots, d$ are assumed to be univariate strict white noise processes with zero mean and unit variance. The set of equations (4.2) simply defines each marginal process as a univariate GARCH. Now we couple the d processes (4.2) imposing a copula family to the multivariate distribution of $\mathbf{Z}_t$. Assume that $\mathbf{Z}_t$ has a d–dimensional copula C with time dependent parameter vector $\boldsymbol{\theta}_t = (\theta_{1,t}, \theta_{2,t}, \ldots, \theta_{e,t})$ such that

$$\theta_{m,t} = r_0 + \sum_{i=1}^{r} r_i \prod_{j=1}^{d} Z_{j,t-i} + \sum_{k=1}^{s} s_k \theta_{m,t-k} \tag{4.3}$$

for $m = 1, 2, \ldots, e$ and where r_i for $i = 0, 1, \ldots, r$ and s_k with $k = 1, 2, \ldots, s$ are scalar model parameters. Equation (4.3) defines a dynamic structure of GARCH type for the dependence parameters and is motivated by (4.1). See also Patton [67] and Rockinger and Jondeau [72] for

related models. Of course, one can attempt to find more suitable dynamics depending on the interpretation that a specific dependence parameter may have. For instance referring to the matrix–diagonal GARCH, the degrees of freedom for the t–innovations were assumed to be constant over time. Asymmetry in the dependence parameters can also be included as for instance for the correlation in the Asymmetric Generalized Dynamic Conditional Correlation GARCH in Cappiello et al. [10].

4.3 Model estimation

The natural estimation method for (4.2) and (4.3) is (conditional) maximum likelihood. Furthermore, the definition of the model suggests a two step estimation procedure. In fact, this is used in similar situations in Engle and Sheppard [30], Patton [67] and Rockinger and Jondeau [72].

Let $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ be a random sample of d–dimensional vectors. Suppose that $\mathbf{X}_t$, given the past information, have continuous marginal distributions F_i parameterised by the parameter vector $\boldsymbol{\alpha}_{i,t}$ for $i = 1, 2, \ldots, d$. Then the multivariate conditional distribution function of $\mathbf{X}_t$ is

$$F(\mathbf{x}; \boldsymbol{\alpha}_{1,t}, \ldots, \boldsymbol{\alpha}_{d,t}, \boldsymbol{\theta}_t) = C(F_1(x_1; \boldsymbol{\alpha}_{1,t}), \ldots, F_d(x_d; \boldsymbol{\alpha}_{d,t}); \boldsymbol{\theta}_t)$$

where C is the copula family of $\mathbf{X}_t$. The conditional density function of $\mathbf{X}_t$ is

$$\begin{aligned} f(\mathbf{x}; \boldsymbol{\alpha}_{1,t}, \ldots, \boldsymbol{\alpha}_{d,t}, \boldsymbol{\theta}_t) &= \\ &= c(F_1(x_1; \boldsymbol{\alpha}_{1,t}), \ldots, F_d(x_d; \boldsymbol{\alpha}_{d,t}); \boldsymbol{\theta}_t) \prod_{i=1}^{d} f_i(x_i; \boldsymbol{\alpha}_{i,t}). \end{aligned}$$

Here we assume that C has a density c given by

$$c(u_1, u_2, \ldots, u_d; \boldsymbol{\theta}) = \frac{\partial^d C(u_1, u_2 \ldots, u_d; \boldsymbol{\theta})}{\partial u_1 \, \partial u_2 \ldots \partial u_d},$$

with $(u_1, u_2, \ldots, u_d) \in [0, 1]^d$ and that F_i has a density f_i for all $i = 1, 2, \ldots, d$. The conditional log–likelihood function of the model then

is

$$\sum_{t=m+1}^{n} \left(\log c(F_1(x_{1,t}; \boldsymbol{\alpha}_{1,t}), \ldots, F_d(x_{d,t}; \boldsymbol{\alpha}_{d,t}); \boldsymbol{\theta}_t) + \sum_{i=1}^{d} \log f_i(x_{i,t}; \boldsymbol{\alpha}_{i,t}) \right) \tag{4.4}$$

where $m = \max(p, r)$. Numerical maximisation of (4.4) gives the maximum likelihood estimates of the model. However, the optimisation of the likelihood function with possibly many parameters is numerically difficult and time consuming. It is more tractable to estimate first the marginal model parameters and then the dependence model parameters using the estimates from the first step. In order to do so, the d marginal likelihood functions

$$\sum_{t=p+1}^{n} \log f_i(x_{i,t}; \boldsymbol{\alpha}_{i,t}) \qquad \text{for } i = 1, 2, \ldots, d, \tag{4.5}$$

are independently maximised. From here we obtain the estimates $\hat{\boldsymbol{\alpha}}_{1,t}, \ldots, \hat{\boldsymbol{\alpha}}_{d,t}$. These are plugged in (4.4) where some terms became constant and can be ignored. The final function to maximise becomes

$$\sum_{t=m+1}^{n} \log c(F_1(x_{1,t}; \hat{\boldsymbol{\alpha}}_{1,t}), \ldots, F_d(x_{d,t}; \hat{\boldsymbol{\alpha}}_{d,t}); \boldsymbol{\theta}_t). \tag{4.6}$$

From this, dependence estimates $\hat{\boldsymbol{\theta}}_t$ are obtained and the model is fitted. In Section 4.4 below we highlight the above procedure, however for space considerations restricting attention to the t–copula.

4.4 Fitting the time–varying copula model to the FX returns

For the USD/DEM and USD/JPY spot rate returns we found that the t–copula yields a good model for the cross dependence. This was shown first through estimating the stationary bivariate distribution (static models); see Chapter 2. In Chapter 3 we arrived at this result though estimating the dependence structure after using time dependent marginal models (fitting copula–models to the filtered returns). In all cases where we fitted

univariate GARCH type models, t–innovations were used. In a next step we combine dynamic models for the margins with dynamic copula models. For each marginal time series we assume a GARCH type model (4.2). For the dependence structure we use a t–copula allowing for time dynamics in the copula parameters (as in (4.3)) which are here the degrees of freedom ν and the correlation ρ:

$$\begin{aligned} \nu_t &= \nu \qquad \text{for all } t, \\ \rho_t &= h^{-1}(r_0 + r_1 z_{1,t-1} z_{2,t-1} + s_1 h(\rho_{t-1})), \end{aligned} \tag{4.7}$$

where $h(\cdot)$ is Fisher's transformation for the correlation

$$h(\rho) = \log\left(\frac{1+\rho}{1-\rho}\right).$$

The choice of model (4.7) is based on tractability and relevance for practice, at the same time it allows us to highlight the general procedure.

Marginal modelling

For every time frequency, each marginal deseasonalised time series is modelled by an ARMA–GARCH model (3.3) with t–innovations. This modelling already took place in Section 3.1. From the univariate fitting we need the estimated degrees of freedom $\hat{\nu}$ to be used in (4.6). The order of the models used was given in Table 3.3. Table 4.2 has the estimated marginal degrees of freedom recalled from Table 3.3.

Dynamic copula modelling results

Denote by $\{(\hat{z}_{1,t}, \hat{z}_{2,t}) : t = 0, \ldots, n\}$ the standardised residual return series obtained from the univariate filtering. The estimated degrees of freedom for the marginal innovation distributions are $\hat{\nu}_1$ and $\hat{\nu}_2$ for USD/DEM and USD/JPY, respectively. Now the standardised residuals are mapped into the unit square by the standard t probability–integral transformation. The probability–integral transformation produces the bivariate time series

Time frequency	$\hat{\nu}$ (s.e.) USD/DEM	USD/JPY
1 hour	3.693 (0.054)	3.654 (0.052)
2 hours	3.708 (0.044)	3.759 (0.077)
4 hours	3.975 (0.105)	3.819 (0.109)
8 hours	4.679 (0.234)	4.357 (0.195)
12 hours	5.385 (0.326)	4.574 (0.251)
1 day	5.797 (0.556)	4.889 (0.412)

Table 4.2: *The degrees of freedom estimated for the marginal conditional distribution t of the innovations and corresponding standard errors.*

of pseudo–observations:

$$\left\{ \left(t_{\hat{\nu}_1} \left(\sqrt{\frac{\hat{\nu}_1}{\hat{\nu}_1 - 2}}\, \hat{z}_{1,t} \right), t_{\hat{\nu}_2} \left(\sqrt{\frac{\hat{\nu}_2}{\hat{\nu}_2 - 2}}\, \hat{z}_{2,t} \right) \right), t = 1, \ldots, n \right\}. \qquad (4.8)$$

This time series, plugged into (4.6) with C being the t–copula function and using (4.7) for the dynamics of the dependence parameters, gives the maximum likelihood estimates for the copula degrees of freedom and correlation (time–varying) parameters.

The results from fitting the dynamic copula model are in Table 4.3. We added the t–copula parameter estimates obtained with no dynamics in the correlation which corresponds to $r_1 = 0$ and $s_1 = 0$ in (4.7) or equivalently, to fit a t–copula with time–invariant parameters to (4.8).

The AIC of the time–varying copula model is lower than the AIC of the constant copula model; see Table 4.3. So we have an improvement in the fitting. The estimate for r_0 can be considered zero for 8 hours, twelve hours and daily returns. But r_1 and s_1 are definitely different from zero for all frequencies. In other words, the estimated (copula) correlation depends on the marginal returns and on the correlation from the previous moment in time. From the estimated parameters for the correlation dynamics we compute, through the second equation of (4.7), the time–varying estimated correlation which is plotted in Figures 4.3 and 4.4. These also show the estimated constant correlation with a 95% confidence interval. Figures 4.1 and 4.2 plot the estimated time–varying

correlation from a matrix–diagonal GARCH model ((4.1)) with bivariate t–innovations. Comparing the two results, the main difference is that the correlation given by the dynamic copula model is much less jagged than the one from the matrix–diagonal GARCH allowing for a more detailed observation of the correlation path.

The number of degrees of freedom estimated for the dependence structure is always larger for the time–varying copula model than for the matrix–diagonal model. In both models the copula used is the t and the margins are t distributed. But while in the matrix–diagonal model margins and copula must have the same degrees of freedom, in the copula–based model they do not. Actually for the daily returns, we can see, from Table 4.2 that for each margin we have $\hat{\nu}_{USD/DEM} = 5.797$, $\hat{\nu}_{USD/JPY} = 4.889$ and from Table 4.3 that for the copula $\hat{\nu} = 8.573$. On the other hand, model (4.1) imposes $\nu_{USD/DEM} = \nu_{USD/JPY} = \nu$ and gives $\hat{\nu} = 5.998$. The degrees of freedom obtained with the matrix–diagonal model are close to those given by the copula time–invariant model; compare the values listed in Tables 4.1 and 4.3. From the different ways the degrees of freedom were estimated there is a common increasing pattern from higher to lower frequencies.

Comments

Whereas in the previous chapters the copula–based models had always a time–invariant dependence structure, this chapter introduced time–varying copula based models. The fitting of matrix–diagonal GARCH type models to the FX returns revealed time variability in the dependence structure. With very few exceptions (see [33], [67] and [72]) multivariate time series models use a multivariate distribution for the innovations. Meaning that, they do not allow to model separately the univariate distributions and the dependence structure. This is exactly the usefulness of copula–based models. We can specify known univariate models for the margins and couple them with a time–varying copula. In this way we

proposed in this chapter such type of models with a particular definition for the parameter dynamics in the case of the t–copula. Assuming t–marginal innovations and a t–copula we fitted the model to the FX returns. We observed that the degrees of freedom of the margins are considerably lower than those of the copula. Hence the marginal FX returns have a much heavier tail than the dependence structure between them. It would be interesting to investigate in a further step t–copula models with time varying degrees of freedom for the returns on the USD/DEM and USD/JPY spot rates. Note that the model is completely general and allows for the combination of a broad spectrum of univariate and copula families.

Time frequency	Parameter Estimates ($\hat{s.e.}$) non–dynamic		dynamic	
1 hour	$\hat{\nu}$	4.935 (0.108)	$\hat{\nu}$	6.330 (0.167)
	$\hat{\rho}$	0.558 (0.002)	$\hat{r}_0$	0.0005 (0.0002)
			$\hat{r}_1$	0.0193 (0.0010)
			$\hat{s}_1$	0.9921 (0.0005)
	AIC	-31517.70	AIC	-34488.72
2 hours	$\hat{\nu}$	4.822 (0.147)	$\hat{\nu}$	6.203 (0.230)
	$\hat{\rho}$	0.580 (0.003)	$\hat{r}_0$	-0.0004 (0.0002)
			$\hat{r}_1$	0.0128 (0.0009)
			$\hat{s}_1$	0.9952 (0.0004)
	AIC	-17192.73	AIC	-19349.29
4 hours	$\hat{\nu}$	4.669 (0.195)	$\hat{\nu}$	6.072 (0.313)
	$\hat{\rho}$	0.592 (0.005)	$\hat{r}_0$	-0.0008 (0.0002)
			$\hat{r}_1$	0.0147 (0.0011)
			$\hat{s}_1$	0.9947 (0.0004)
	AIC	-9085.848	AIC	-10262.23
8 hours	$\hat{\nu}$	5.296 (0.339)	$\hat{\nu}$	7.206 (0.584)
	$\hat{\rho}$	0.612 (0.006)	$\hat{r}_0$	0.0005 (0.0005)
			$\hat{r}_1$	0.0173 (0.0014)
			$\hat{s}_1$	0.9927 (0.0006)
	AIC	-4813.6	AIC	-5456.312
12 hours	$\hat{\nu}$	5.830 (0.499)	$\hat{\nu}$	8.053 (0.884)
	$\hat{\rho}$	0.620 (0.008)	$\hat{r}_0$	0.0002 (0.0008)
			$\hat{r}_1$	-0.0249 (0.0023)
			$\hat{s}_1$	0.9901 (0.0010)
	AIC	-3299.16	AIC	-3744.28
1 day	$\hat{\nu}$	5.945 (0.758)	$\hat{\nu}$	8.573 (1.455)
	$\hat{\rho}$	0.619 (0.011)	$\hat{r}_0$	-0.0023 (0.0017)
			$\hat{r}_1$	-0.0343 (0.0041)
			$\hat{s}_1$	0.9846 (0.0021)
	AIC	-1644.549	AIC	-1881.760

Table 4.3: *Parameter estimates, standard errors and AIC values for the two copula models, without and with dynamics in the correlation, fitted to the hourly up to daily returns on USD/DEM and USD/JPY rates.*

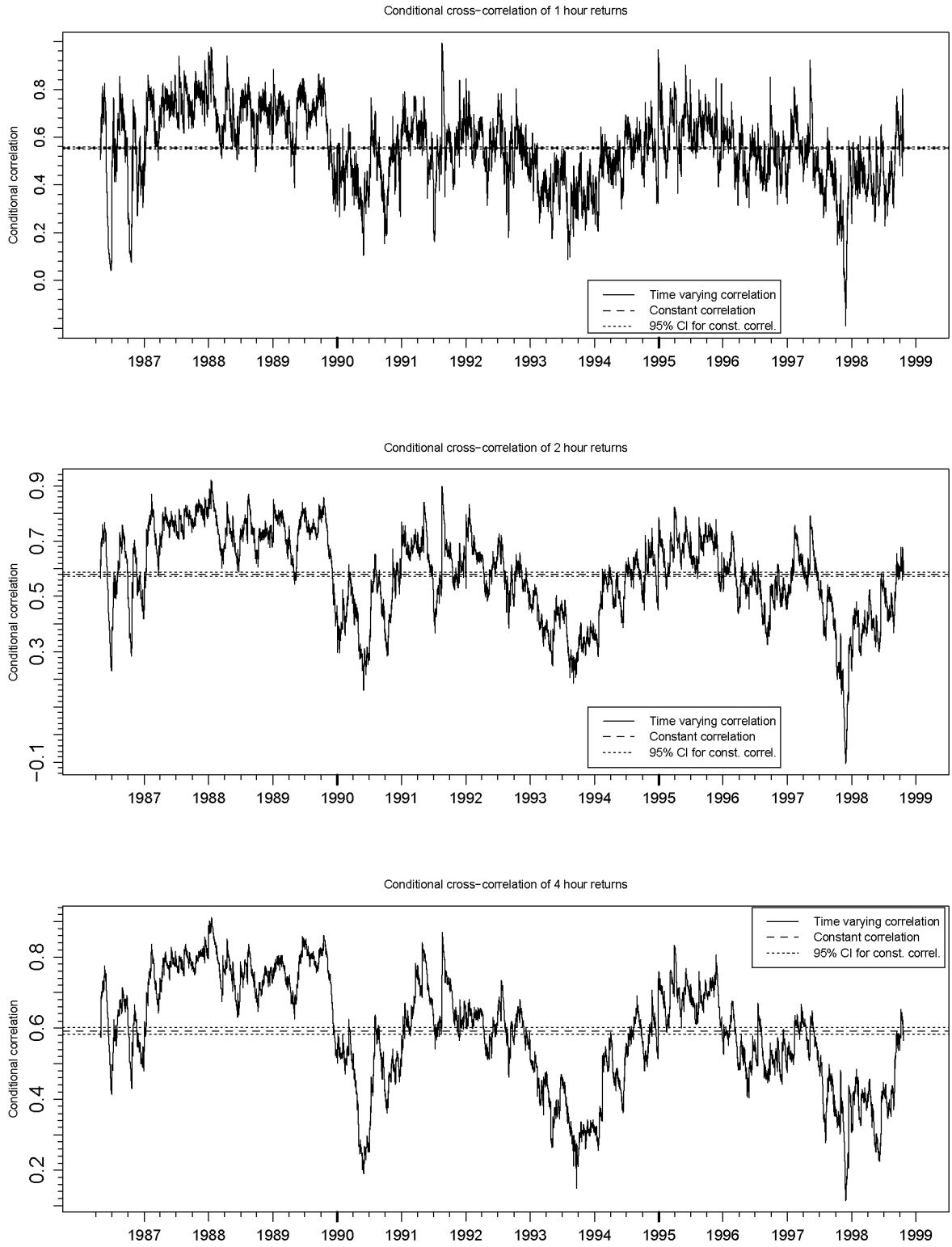

Figure 4.3: *Time–varying cross–correlations estimated by a time–varying copula–based model for the one, two and four hours returns on the FX USD/DEM and USD/JPY spot rates.*

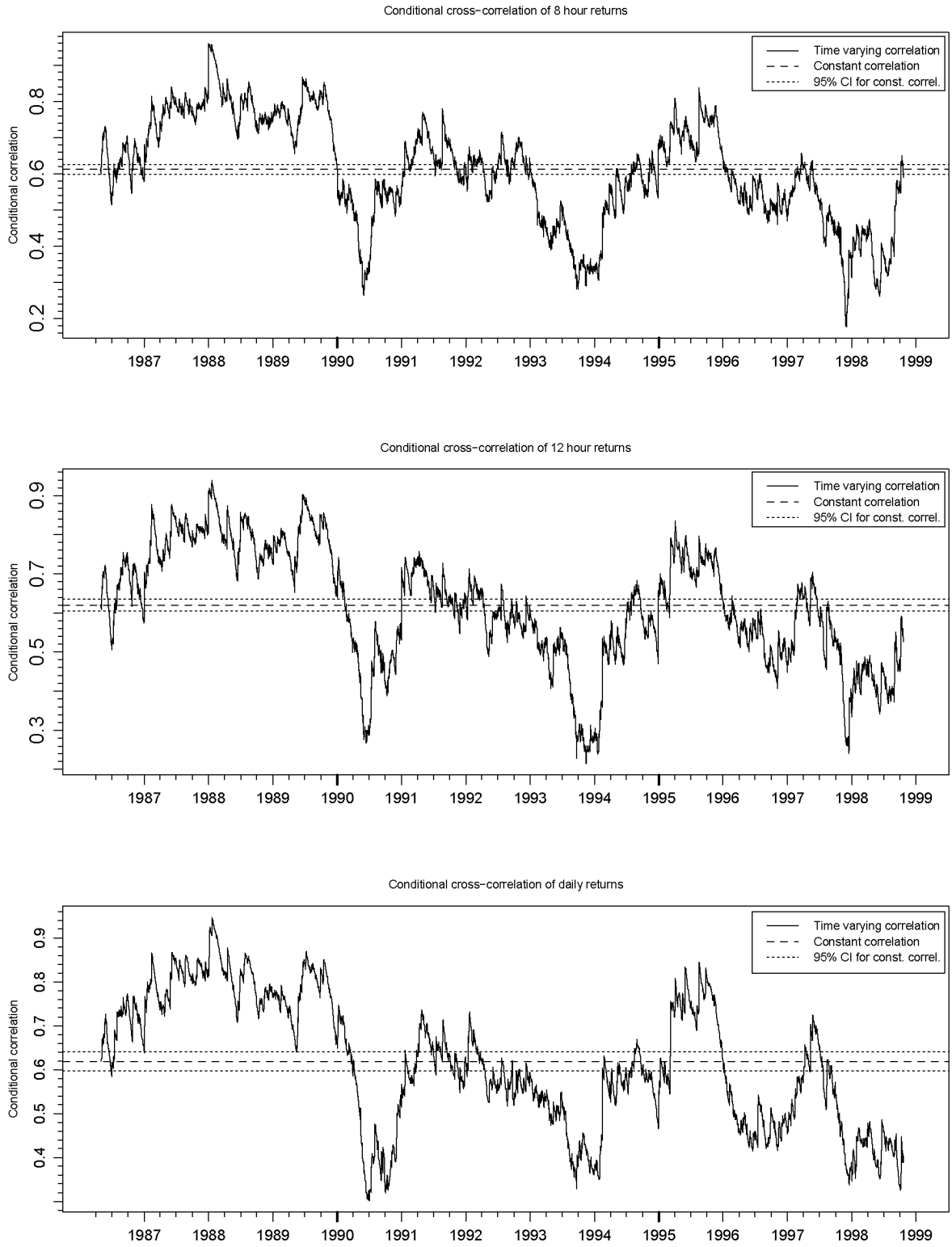

Figure 4.4: *Time–varying cross–correlations estimated by a time–varying copula–based model for the eight hours, twelve hours and daily returns on the FX USD/DEM and USD/JPY spot rates.*

Chapter 5

Change–point analysis for copulae

In Chapter 3 we modelled multivariate time series using for the dynamics of each margin a GARCH type model and for the contemporaneous dependence a time–invariant copula model. If the data cover a reasonably long period of time, more than one year say, we may expect that economic factors induce changes if not in the copula family at least in the dependence parameters. Conceptually in the other extreme, in Chapter 4 we modelled dynamically the copula parameters allowing these to change "continuously" in time. Because these models are very versatile, they are more difficult to fit. Meaning that we have to be able to specify the dynamics of the copula parameters very close to "reality" in order to have good results. In this chapter we consider the specific sub–problem of change–point detection for parametric copula models. For some related work, see for instance Gombay and Horváth [41]. For a detailed treatment of the change–point theory underlying our approach, see Csörgő and Horváth [14] and references therein.

The chapter is organised as follows: in Section 5.1 we summarise the max-

imum likelihood approach of change–point analysis with the detection of changes in the dependence structure in mind. We compute the distribution of the resulting test statistic for some commonly used parametric copulae, infer about the power of the test, discuss the construction of confidence intervals and illustrate the methods introduced on a simulated data example. Section 5.2 has an analysis for two financial positions on the same simulated data, using the methodologies presented in the previous section. An application of the methods introduced on real insurance data is given in Section 5.3. In Section 5.4 we test for change–points in the dependence structure of the FX returns on USD/DEM and USD/JPY data set.

5.1 Statistical change–point analysis

5.1.1 The test statistic

Suppose that we have n vectors of observations, each composed by d risks. Formally, let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a sequence of independent random vectors in $\mathbb{R}^d$ with distribution functions $F(\mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\eta}_1), \ldots, F(\mathbf{x}; \boldsymbol{\theta}_n, \boldsymbol{\eta}_n)$, respectively, where $\boldsymbol{\theta}_i$ and $\boldsymbol{\eta}_i$ are parameters of the distribution functions such that $\boldsymbol{\theta}_i \in \Theta^{(1)} \subseteq \mathbb{R}^q$ and $\boldsymbol{\eta}_i \in \Theta^{(2)} \subseteq \mathbb{R}^p$ for $1 \leq i \leq n$. We will be primarily interested in a change–point analysis for the $\boldsymbol{\theta}_i$'s, whereas the $\boldsymbol{\eta}_i$'s will be nuisance parameters. As a consequence, we are interested in testing the null hypothesis

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \ldots = \boldsymbol{\theta}_n \qquad \text{and} \qquad \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = \ldots = \boldsymbol{\eta}_n$$

versus the alternative

$$H_A : \boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_{k^*} \neq \boldsymbol{\theta}_{k^*+1} = \ldots = \boldsymbol{\theta}_n \quad \text{and} \quad \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = \ldots = \boldsymbol{\eta}_n.$$

Here k^* is the location or time of the (assumed) single change–point. All the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta}) \in \Theta^{(1)} \times \Theta^{(2)}$ are supposed to be unknown under both hypotheses. As a start, assume that $k^* = k$ is known. In that

case, the question consists of testing whether two samples come from the same population and can be achieved through the generalised likelihood ratio test. The null hypothesis will be rejected for small values of the test statistic

$$\Lambda_k = \frac{\sup_{(\boldsymbol{\theta},\boldsymbol{\eta})\in\Theta^{(1)}\times\Theta^{(2)}} \prod_{1\le i\le n} f(\mathbf{X}_i;\boldsymbol{\theta},\boldsymbol{\eta})}{\sup_{(\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\eta})\in\Theta^{(1)}\times\Theta^{(1)}\times\Theta^{(2)}} \prod_{1\le i\le k} f(\mathbf{X}_i;\boldsymbol{\theta},\boldsymbol{\eta}) \prod_{k<i\le n} f(\mathbf{X}_i;\boldsymbol{\theta}',\boldsymbol{\eta})}. \quad (5.1)$$

As the estimation of Λ_k is carried out through maximum likelihood, all the necessary conditions of regularity and efficiency have to be assumed (see for instance [54]).

If we denote

$$L_k(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{1\le i\le k} \log f(\mathbf{X}_i;\boldsymbol{\theta},\boldsymbol{\eta})$$

and

$$L_k^*(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{k<i\le n} \log f(\mathbf{X}_i;\boldsymbol{\theta},\boldsymbol{\eta})$$

then the likelihood ratio equation (5.1) can be written as

$$-2\log(\Lambda_k) = 2[L_k(\hat{\boldsymbol{\theta}}_k,\hat{\boldsymbol{\eta}}_k) + L_k^*(\boldsymbol{\theta}_k^*,\hat{\boldsymbol{\eta}}_k) - L_n(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\eta}}_n)].$$

As stated, the statistic Λ_k tests H_0 for k known. In the (more realistic) case when k is unknown, H_0 will be rejected for large values of

$$Z_n = \max_{1\le k<n} (-2\log(\Lambda_k)). \quad (5.2)$$

The asymptotic distribution of $Z_n^{1/2}$ can be derived using Extreme Value Theory (EVT) techniques; see [24] and [27]. Indeed, let

$$A(x) = (2\log(x))^{1/2}$$

and

$$D_q(x) = 2\log(x) + \frac{q}{2}\log(\log(x)) - \log(\Gamma(q/2)),$$

where $\Gamma(t)$ is the gamma function

$$\Gamma(t) = \int_0^\infty y^{t-1} \exp(-y)\, dy.$$

Then, if H_0 and all the necessary regularity conditions hold, we have that

$$\lim_{n\to\infty} P(A(\log(n))Z_n^{1/2} \leq t + D_q(\log(n))) = \exp(-2\exp(-t)) \qquad (5.3)$$

for all t. The right hand side of (5.3) is the square of a Gumbel distribution function. It is known that the rate of convergence in results like (5.3) is very slow; see for instance Embrechts et al. [27, page 150]. Therefore care has to be taken to base a hypothesis test on the asymptotic distribution, especially for small and moderate sample sizes. For this reason Gombay and Horváth [40] derived a result on which a test can be based that yields better rejection regions for smaller sample sizes. Under H_0 and supposing that all the necessary regularity conditions hold, for $n \to \infty$,

$$\left| Z_n^{1/2} - \sup_{1/n\leq t\leq 1-1/n} \left(\frac{B_n^{(q)}(t)}{t(1-t)} \right)^{1/2} \right| = o_P(\exp(-(\log(n))^{1-\varepsilon})) \qquad (5.4)$$

for all $0 < \varepsilon < 1$, where $\{B_n^{(q)} : 0 \leq t \leq 1\}$ is a sequence of stochastic processes such that $\{B_n^{(q)} : 0 \leq t \leq 1\} \stackrel{d}{=} \{B^{(q)} : 0 \leq t \leq 1\}$ for each n and $B^{(q)}(t) = \sum_{1\leq i\leq q} B_i^2(t)$, where $\{B_s(t) : 0 \leq t \leq 1\}$, $s = 1, \ldots, q$ are independent Brownian bridges.

For $0 < \alpha < 1$ let

$$z_n = z_n(1-\alpha) = \sup\left\{ x \geq 0 : P(Z_n^{1/2} \leq x) \leq 1-\alpha \right\}$$

and

$$\begin{aligned} u(h,l) &= u(h,l;1-\alpha) \\ &= \sup\left\{ x \geq 0 : P\left(\sup_{h\leq t\leq 1-l} \left\{ \frac{B^{(q)}(t)}{t(1-t)} \right\}^{1/2} \leq x \right) \leq 1-\alpha \right\}. \end{aligned}$$

As a consequence of (5.3) and (5.4), if $h(n) \geq 1/n$, $l(n) \geq 1/n$ and for $n \to \infty$,

$$(2\log\log n)^{-1/2} \sup_{1-c(n)\leq t\leq 1-1/n} \left\{ \frac{B_n^{(q)}(t)}{t(1-t)} \right\} \xrightarrow{P} 1-\varepsilon^*$$

is satisfied for some $0 < \varepsilon^* \leq 1$, where

$$c(n) = \exp((\log n)^{1-\varepsilon^*})/n,$$

we have that

$$\lim_{n\to\infty} P(Z_n^{1/2} > u(h(n), l(n))) = \alpha$$

and

$$|z_n(1-\alpha) - u(h(n), l(n))| = o((\log\log n)^{-1/2});$$

see [14, page 24]. The distribution function of $\sup_{h\leq t\leq l}\{B^{(q)}(t)/(t(1-t))\}^{1/2}$ has no simple closed form. For practical applications we use, for $0 < h < l < 1$, the following approximation:

$$P\left(\sup_{h\leq t\leq 1-l}\left\{\frac{B^{(q)}(t)}{t(1-t)}\right\}^{1/2} \geq x\right) = \frac{x^q \exp(-x^2/2)}{2^{q/2}\Gamma(q/2)}\cdot$$
$$\cdot\left(\left(1-\frac{q}{x^2}\right)\log\frac{(1-h)(1-l)}{hl} + \frac{4}{x^2} + O\left(\frac{1}{x^4}\right)\right), \quad (5.5)$$

as $x \to \infty$; see [40].

Based on (5.3) and (5.4) asymptotic critical values can be computed. Table 5.1 has the values computed from (5.3) in column $z^{(1)}$ and from (5.4) in column $z^{(2)}$. Gombay and Horváth [40] found that $h(n) = l(n) = (\log n)^{3/2}/n$ makes $u(h, l)$ a good approximation for $z_n = z_n(1-\alpha)$. We use the same choice to obtain the asymptotic critical values $z^{(2)}$ listed in Table 5.1. On the other hand, for a given model, we can perform Monte Carlo simulations in order to get a further approximation to the critical values of the likelihood ratio test statistic $Z_n^{1/2}$. Under the null hypothesis, $N = 5\,000$ repetitions of $Z_n^{1/2}$ for $n = 50$, $n = 100$ and $n = 500$ were performed in the case of bivariate Gumbel, Frank and Gaussian copulae. If we denote the simulated repetitions of $Z_n^{1/2}$ by $Z_{n,t}^{1/2^*}$, $t = 1, \ldots, N$, then the estimated critical value z_n^* at the level $1-\alpha$ is given by $Z_{n,((1-\alpha)(N+1))}^{1/2^*}$ where $Z_{n,(r)}^{1/2^*}$ is the rth ordered value (For an explanation on Monte Carlo quantiles, see for instance [42]). These results are also to be found in Table 5.1.

Table 5.1 can be used to test whether the parameter of the Gumbel or the Frank copulae changed for a given data set. Or we can test whether the

Sample size	$1-\alpha$	$z^{(1)}$	$z^{(2)}$	z_n^{Gu}	z_n^{Fr}	z_n^{Ga}
50	0.90	3.18	2.69	2.67	2.51	2.87
	0.95	3.62	2.97	2.93	2.76	3.13
	0.99	4.60	3.52	3.45	3.22	3.59
100	0.90	3.23	2.79	2.81	2.64	2.95
	0.95	3.64	3.06	3.02	2.88	3.19
	0.99	4.57	3.59	3.54	3.39	3.78
500	0.90	3.31	2.95	2.94	2.89	3.07
	0.95	3.69	3.20	3.18	3.13	3.32
	0.99	4.54	3.71	3.61	3.68	3.79

Table 5.1: *Critical values for the likelihood ratio test $Z_n^{1/2}$ given by (5.2). The third column has the asymptotic critical values given by (5.3), in the fourth column the $z^{(2)}$'s are derived from (5.4). The z_n^{Gu}, z_n^{Fr} and z_n^{Ga} are the simulated critical values for the Gumbel, Frank and Gaussian bivariate copulae, respectively. The simulations are made under H_0 and one parameter can change under the alternative hypothesis ($q = 1$).*

correlation in a Gaussian copula is constant. In Csörgő and Horváth [14] the asymptotic values $z^{(1)}$ and $z^{(2)}$ are compared with univariate normal, exponential and Poisson observations. In Table 5.1 we compare the same asymptotic critical values with those obtained from bivariate Gumbel, Frank and Gaussian copula observations. The simulations show that in fact the true critical values are smaller than the ones derived from (5.3) meaning that the later result usually provides a conservative rejection region. In the case of univariate normal, exponential and Poisson observations, this fact was already observed by Csörgő and Horváth in [14]. In Figure 5.1 we can see that, for the Gumbel case with sample size 100, the asymptotic result (5.4) is much closer to the distribution function of the test statistic than (5.3). In the right panel are the values of the same distribution functions for more than 0.9 of probability. The asymptotic result (5.4) still gives a good approximation in the tail. For the three sample sizes in Table 1 and for the Gumbel and the Frank distributions, we can observe in Figure 5.2 the asymptotic and the simulated distribution functions of $Z_n^{1/2}$. In these plots we see that the asymptotic result (5.4) actually gives a good (small sample) approximation for the rejection

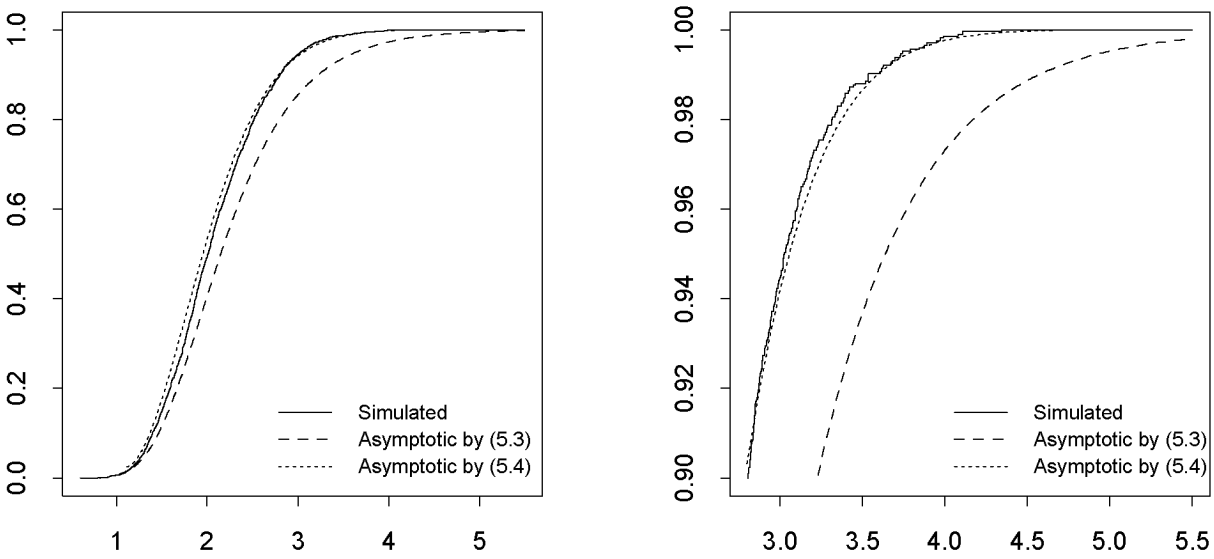

Figure 5.1: *The distribution function of* $Z_{100}^{1/2}$ *given by the asymptotic approximations (5.3) and (5.4) and by simulation for the Gumbel case. The left panel presents the full distribution functions whereas the right panel concentrates on the tail region above 90%.*

regions and is much better than the approximation given by (5.3), for all cases considered.

5.1.2 An example: the Gumbel case

In this chapter we use the Gumbel copula given in Example 1.6 reparameterised. Instead of considering the parameter $\theta \in [1, +\infty)$ we use here $\theta \in (0, 1]$. In this way, it is easier to judge the impact of a parameter change in the degree of dependence. A set of 500 observations was simulated from a bivariate Gumbel copula with parameter $\theta = 0.8$ for the first 250 observations and $\theta = 0.4$ for the second half. Recall from Example 1.6,

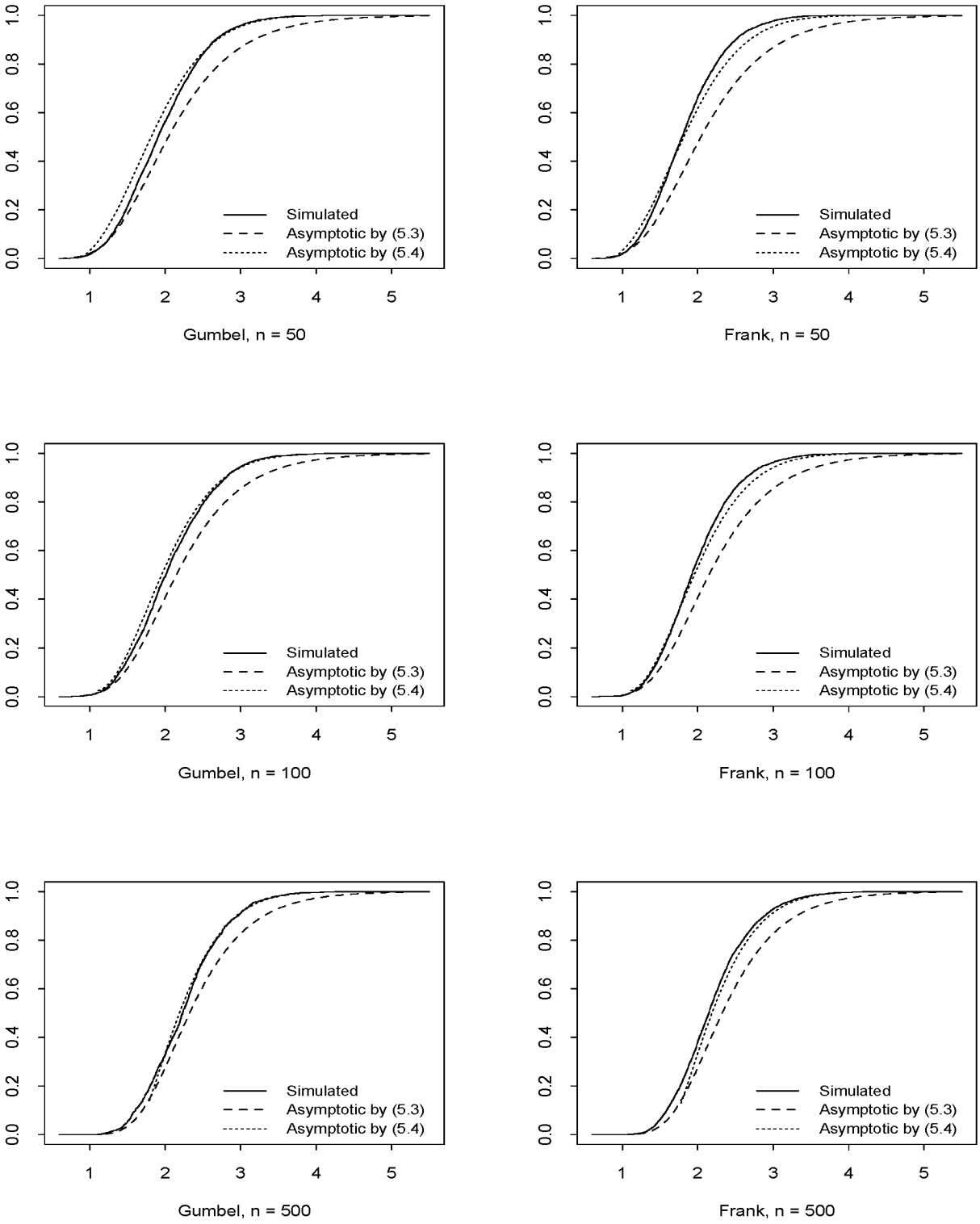

Figure 5.2: *The distribution function of* $Z_n^{1/2}$ *given by the asymptotic results (5.3) and (5.4) and by simulation. The left panels correspond to the Gumbel distribution whereas the right ones to the Frank. The rows of the graphics correspond to different sample sizes* $n = 50$, $n = 100$ *and* $n = 500$.

after substituting θ by $1/\theta$ that a random vector (U, V) has a Gumbel copula if for $0 \leq u, v \leq 1$,

$$P(U \leq u, V \leq v) = \exp\left\{-\left[(-\ln u)^{1/\theta} + (-\ln v)^{1/\theta}\right]^{\theta}\right\}, \quad \theta \in (0, 1].$$

In the case of the Gumbel copula, there is a straightforward link between the parameter θ and the Kendall τ correlation coefficient,

$$\tau = 1 - \theta,$$

hence this simulation can also be interpreted as testing for a change in τ from 0.2 to 0.6. Questions of the latter type are important in various applications; see for instance [41]. We can at first assume that all the data come from a homogeneous model, following a Gumbel copula with unknown parameter θ, say. Fitting this model by maximum likelihood produced an estimate for the copula parameter of $\hat{\theta} = 0.6325$. Figure 5.3 has the scatter plot of the simulated data and a QQ–plot of the fitted against the theoretical quantiles. The QQ–plot points to a fairly good fit, as the pairs of quantiles lay almost all very close to the main diagonal. Actually, we can test whether $V|U$ has indeed a standard uniform distribution. A Kolmogorov–Smirnov goodness of fit test does not reject the fitted Gumbel model (p–value of 0.22). This points to an important (robustness type) property of copula modeling: through a Gumbel model for instance, a whole range of linear correlation values in agreement with the data can be covered.

We now turn to the main issue underlying the example, the performance of the test statistic Λ_k towards the change–point detection from $\theta = 0.8$ to 0.4 around the middle of the data set. Figure 5.4 is the plot of $(-2\log(\Lambda_k))^{1/2}$ for $1 \leq k < n$. The maximum of these values is $z^{1/2}_{500\,obs} = 10.02$, reached at $k = 248$. Either by approximations (5.3), (5.4) or by simulation, $P(Z^{1/2}_{500} \leq 10.02) = 1$, hence we reject H_0 very significantly. Hence our analysis yields a change–point at observation $k = 248$ with pre– and post–values of θ estimated as $\hat{\theta}_b = 0.8515$, $\hat{\theta}_a = 0.4087$, respectively.

The excellent performance of the test in detecting the simulated change in the example above has to be taken with caution. The analysis based on the

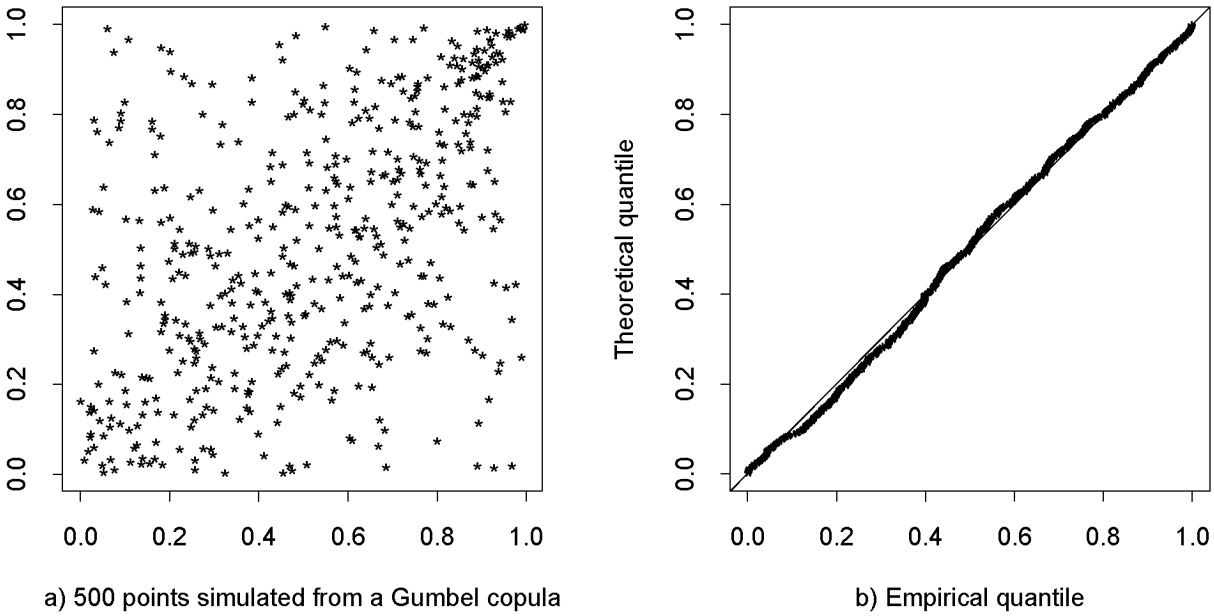

Figure 5.3: *Simulated values from a Gumbel copula with the θ–values 0.8 and 0.4 and QQ–plot corresponding to the maximum likelihood fit, where $\hat{\theta} = 0.6325$.*

assumption of homogeneity (no θ–change) already yields a warning sign. It turns out that the power function for $Z_n^{1/2}$ depends on the location of the change and on the size of the change. For instance, for a small change near the limits of the sampling time interval, the test does not give such good results. The next section is devoted to this issue. Furthermore, in most practical applications, interdependence between the data may further weaken the analysis.

5.1.3 The power of the test

As already discussed above in the Gumbel example, it will be important to analyse the power of the change–point test presented. The power function is defined as follows $\beta(\alpha) = P(Z_n^{1/2} > z_n(\alpha)|H_A)$ where H_A stands for the alternative hypothesis of one change–point. Hence we need the distribution function of the test statistic $Z_n^{1/2}$ under the alternative hypothesis; in the case of interest, i.e. the bivariate Gumbel and Frank copulae, this distribution is unknown. We therefore perform Monte Carlo

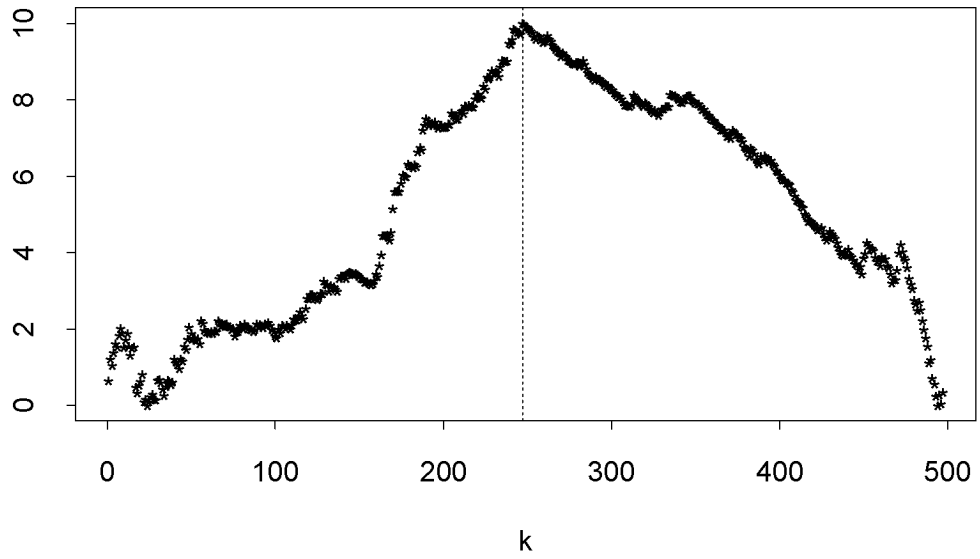

Figure 5.4: *Values of* $(-2\log(\Lambda_k))^{1/2}$, $1 \leq k < 500$, *for the simulated data in the Gumbel example of Section 5.1.2.*

simulations to estimate the power function and replicate $Z_n^{1/2}$, under the alternative, $N = 5\,000$ times for the Gumbel copula and a sample size of $n = 100$. The power function, for a given level $1 - \alpha$, is estimated by $\hat{\beta}_n(\alpha) = (1 + \#\{Z_n^{1/2^*} > z_n(\alpha)\})/(1 + N)$. As we know that the distribution function of $Z_n^{1/2}$ depends also on the size of the change and on the location of the change, under the alternative hypothesis, we simulated several different scenarios. For the location of the change k^*, we took values $10, 25$ and 50. As for the size of the change, we considered the Gumbel distribution with parameter $0 < \theta \leq 1$ and changes between 0.1 and 0.9. The results reported in Table 5.2 are based on the critical values given in Table 5.1 which were obtained by the asymptotic distribution in (5.4).

The values in Table 5.2 confirm that the power of the test strongly depends on the size of the change and on its location. As is to be expected, the bigger the size of the change, the more powerful the test is. The closer the change–point k^* comes to the edge of the data, the less powerful the test becomes. Based on Table 5.2, it seems that the loss of power is more due to a size change than to the location of the change. For applications this is relevant as it is more important to get an early warning once a change

	Level	Size of the change			
k^*	$1-\alpha$	0.1	0.3	0.5	0.9
	0.90	0.1330	0.4028	0.8416	1.000
10	0.95	0.0618	0.2642	0.7341	1.000
	0.99	0.0130	0.0890	0.4423	1.000
	0.90	0.1748	0.7314	0.9984	1.000
25	0.95	0.0932	0.6092	0.9928	1.000
	0.99	0.0202	0.3466	0.9596	1.000
	0.90	0.2030	0.8650	0.9998	1.000
50	0.95	0.1140	0.7902	0.9994	1.000
	0.99	0.0290	0.5846	0.9970	1.000

Table 5.2: *Power function values of the change–point test for different locations k^* and sizes of the change. The sample size is $n = 100$. Each value was obtained from 5 000 simulations of the corresponding case.*

has taken place.

So far we have tested for the existence of a change–point. In the next section we will look more in detail at the estimation of the time of a change–point, also constructing confidence intervals for this time.

5.1.4 The time of the change and corresponding confidence intervals

If we assume that there is exactly one change–point, then the maximum likelihood estimator for the time of the change is given by

$$\hat{k}_n = \min\{1 \leq k < n : Z_n = -2\log(\Lambda_k)\}. \tag{5.6}$$

In the case that there is no change, $\hat{k}_n$ will take a value near the limits of the sample. This holds because under the null hypothesis and if all the necessary regularity conditions hold, for $n \to \infty$,

$$\hat{k}_n/n \xrightarrow{d} \xi, \tag{5.7}$$

where $P(\xi = 0) = P(\xi = 1) = 1/2$; see [14, page 51]. Figure 5.5 gives the frequency plot of $\max_{1\leq k<500}(-2\log(\Lambda_k))$ for 5 000 samples of size

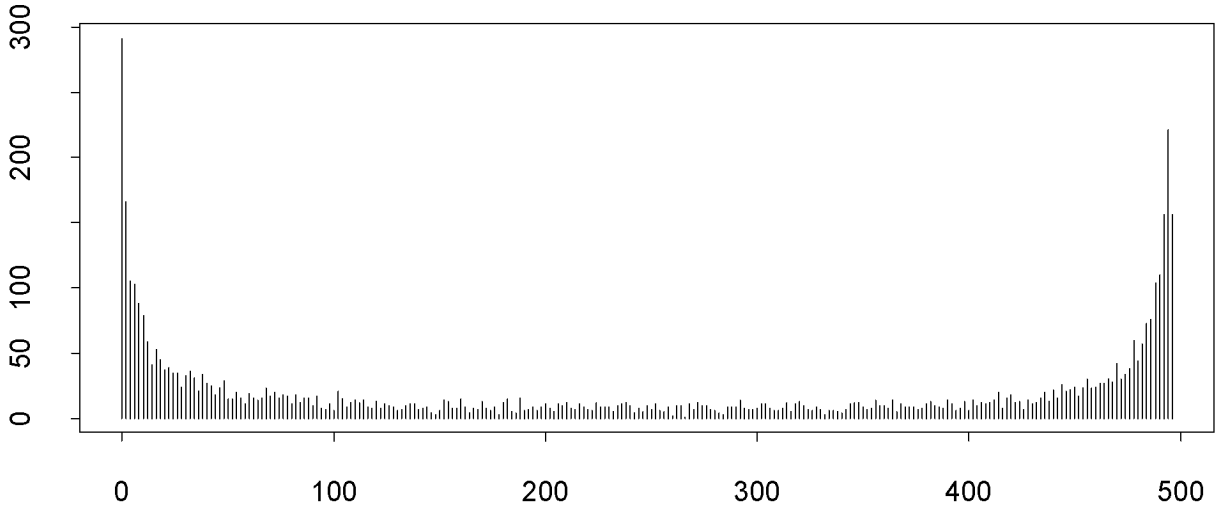

Figure 5.5: *Frequency plot of the time k for which $Z_{500} = \max_{1 \le k < 500} (-2 \log(\Lambda_k))$ is attained, based on 5 000 simulations of Gumbel copula samples of size $n = 500$ with $\theta = 0.5$ (no change–point).*

500 simulated from a Gumbel copula, with parameter $\theta = 0.5$, under the null hypothesis of no change. We can see from that graph that indeed the estimator of the time of the change mostly lays near the beginning or near the end of the time scale.

The value $\hat{k}_n$ in (5.6) gives an estimate for the time of the change. In order to construct a confidence interval for the time of the change, we need to know or to approximate the distribution of $\hat{k}_n - k_0$, where k_0 is the true time of the change. One approximation can be obtained using bootstrap methodology. In the literature, there are several different approaches to construct bootstrap confidence intervals; see for instance [17], [23] or [42]. One of the simplest ways for constructing such an interval is the percentile method. A theoretical $(1 - 2\alpha)100\%$ confidence interval for the time of the change k_0 is of the form

$$(\hat{k}_n - a_2, \hat{k}_n - a_1) \tag{5.8}$$

where a_1 and a_2 satisfy

$$P(\hat{k}_n - a_2 \le k_0 \le \hat{k}_n - a_1) = 1 - 2\alpha.$$

Moreover if we want an equi–tailed interval, then we require that

$$P(k_0 \leq \hat{k}_n - a_2) = \alpha = P(k_0 > \hat{k}_n - a_1), \tag{5.9}$$

hence a_1 and a_2 are quantiles of the random variable $\hat{k}_n - k_0$.

Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is the sample from which we estimated the time of the change $\hat{k}_n$ assuming that the data came from a population with distribution function in the copula family C_θ. Then we have the maximum likelihood estimates $\hat{\theta}_b$ and $\hat{\theta}_a$ of the parameter of the distribution before and after the time of the change, respectively. From this we replicate the original sample simulating N samples of size n from the fitted distribution

$$F_{\hat{\theta}_b, \hat{\theta}_a, \hat{k}_n}(\mathbf{x}_i) = C_{\hat{\theta}_b}(\mathbf{x}_i)\mathbb{I}_{\{j \in \mathbb{N}: j \leq \hat{k}_n\}}(i) + C_{\hat{\theta}_a}(\mathbf{x}_i)\mathbb{I}_{\{j \in \mathbb{N}: j > \hat{k}_n\}}(i)$$

and compute the estimated time of the change $\hat{k}^*_{n,i}$ for each replicate sample, $i = 1, \ldots, N$. These replicates allow us to estimate the distribution function of $\hat{k}^*_n - \hat{k}_n$, where $\hat{k}^*_n$ is the time of the change estimate of a resample from a population with distribution function $F_{\hat{\theta}_b, \hat{\theta}_a, \hat{k}_n}$. Suppose that k^*_α and $k^*_{1-\alpha}$ are the quantiles of $\hat{k}^*_n$ such that

$$P(\hat{k}^*_n \leq k^*_\alpha) = \alpha = P(\hat{k}^*_n > k^*_{1-\alpha}),$$

then

$$P(k^*_\alpha - \hat{k}_n \leq \hat{k}^*_n - \hat{k}_n \leq k^*_{1-\alpha} - \hat{k}_n) = 1 - 2\alpha. \tag{5.10}$$

The bootstrap principle for confidence intervals consists of assuming that we can approximate the quantiles of $\hat{k}_n - k_0$ by the quantiles of $\hat{k}^*_n - \hat{k}_n$. From (5.9) and (5.10) we then have that

$$P(k^*_\alpha - \hat{k}_n \leq \hat{k}_n - k_0 \leq k^*_{1-\alpha} - \hat{k}_n) \approx 1 - 2\alpha,$$

or that

$$P(\hat{k}_n - (k^*_{1-\alpha} - \hat{k}_n) \leq k_0 \leq \hat{k}_n - (k^*_\alpha - \hat{k}_n)) \approx 1 - 2\alpha \tag{5.11}$$

where $(k^*_{1-\alpha} - \hat{k}_n)$ and $(k^*_\alpha - \hat{k}_n)$ are the bootstrap approximations of respectively a_2 and a_1 in (5.8) for the equi–tailed case.

It is well known that the estimation of confidence limits for a transformation of a parameter may give much better results than the direct estimation for the parameter; see for instance [23, pages 54 and 162]. Suppose that there exists a transformation of the random variable $\hat{k}_n$, say $\hat{u}_n = h(\hat{k}_n)$, which has a symmetric distribution. Then applying (5.11) we have that

$$P(\hat{u}_n - (u^*_{1-\alpha} - \hat{u}_n) \leq h(k_0) \leq \hat{u}_n - (u^*_\alpha - \hat{u}_n)) \approx 1 - 2\alpha, \tag{5.12}$$

where u^*_α is the α–quantile of the transformed random variable $h(\hat{k}^*_n)$. Because of symmetry we know that $a_1 = -a_2$ in (5.8). So we can transform the approximated quantiles of $\hat{u}_n - h(k_0)$ in (5.12) such that

$$P(\hat{u}_n - (\hat{u}_n - u^*_\alpha) \leq h(k_0) \leq \hat{u}_n - (\hat{u}_n - u^*_{1-\alpha})) \approx 1 - 2\alpha.$$

Simplifying this expression and transforming back to the original scale we obtain

$$P(k^*_\alpha \leq k_0 \leq k^*_{1-\alpha}) \approx 1 - 2\alpha. \tag{5.13}$$

In order to obtain the quantiles k^*_α from the bootstrap replicates $\hat{k}^*_{n,i}$ $(i = 1, \ldots, N)$ we note that, if $\hat{k}^*_{n,1}, \ldots, \hat{k}^*_{n,N}$ are independent and identically distributed with distribution function F, then

$$E\left(\hat{k}^*_{n,(j)}\right) = F^{-1}\left(\frac{j}{N+1}\right),$$

where, as before, $\hat{k}^*_{n,(j)}$ denotes the jth ordered value. Using this result an estimate for $k^*_\alpha = F^{-1}(\alpha)$ is $\hat{k}^*_{n,((N+1)\alpha)}$. Finally substituting these values in (5.13) we obtain that the bootstrap confidence interval for the time of the change k_0 becomes

$$\left(\hat{k}^*_{n,((N+1)\alpha)}, \hat{k}^*_{n,((N+1)(1-\alpha))}\right). \tag{5.14}$$

The interval in (5.14) is often referred to as the bootstrap percentile confidence interval. Below, we highlight this construction on the example presented in Section 5.1.2.

Example 5.1 In our simulated example, the estimate for the time of the change is $\hat{k}_{500} = 248$; see Figure 5.4. The maximum likelihood parameter

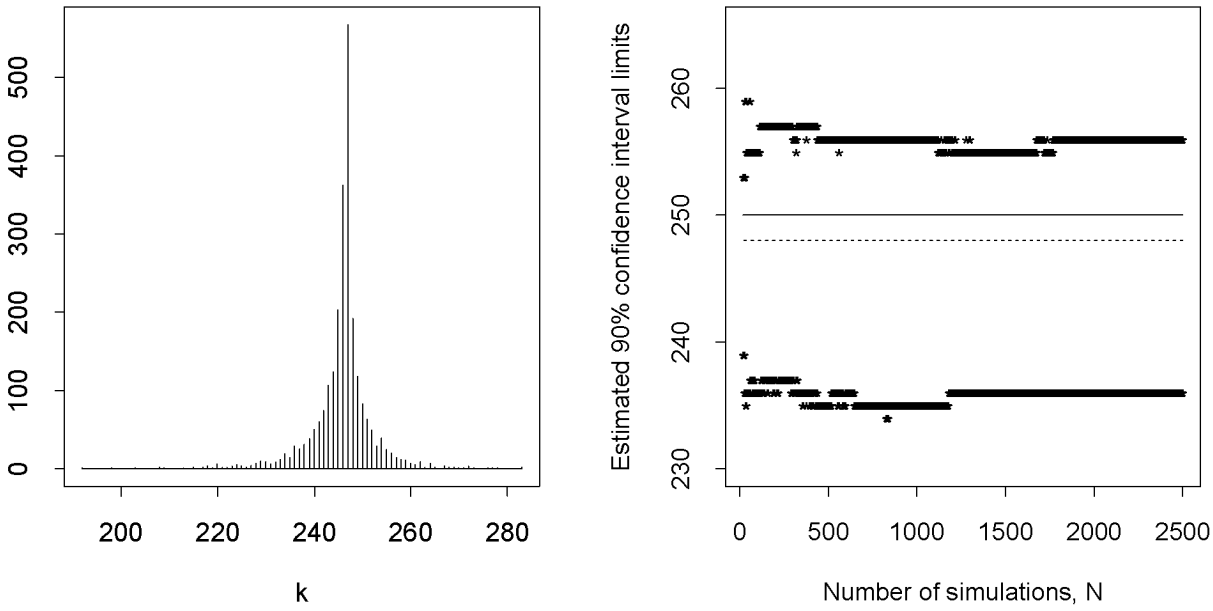

Figure 5.6: *Left panel: Frequency plot of the time of the change estimated from the simulated sets from the fitted distribution (5.15). Right panel:* 90% *bootstrap percentile confidence interval limits under resampling from the fitted model with one change for the data in the example from Section 5.1.2.*

estimates before and after the estimated time of the change are $\hat{\theta}_b = 0.8515$ and $\hat{\theta}_a = 0.4087$, respectively. To obtain a confidence interval for the time of the change we have to simulate N samples of size 500 from the fitted bivariate Gumbel copula with a change in $k = 248$. Each sample is generated from

$$C^{Gu}_{\theta=0.8515}(\mathbf{x}_i)\mathbb{I}_{\{j\in\mathbb{N}:j\leq 248\}}(i) + C^{Gu}_{\theta=0.4087}(\mathbf{x}_i)\mathbb{I}_{\{j\in\mathbb{N}:j>248\}}(i), \qquad (5.15)$$

with $i = 1, \ldots, 500$. The frequency plot of the time of the changes estimated in the N simulations is given in the left panel of Figure 5.6.

In order to get a feeling for the number of bootstrap replicates needed to obtain a stable interval, in Figure 5.6 we plot the interval boundaries as a function of the number of replicates. In that plot the thin filled line is the change time in the data of the example ($k = 250$) and the dotted line is the estimated time of the change ($\hat{k}_{500} = 248$). We see that from around $N = 750$, the interval boundaries seem to stabilise around two values. We

have that the 90% bootstrap percentile confidence interval for the time of the change given by (5.14) based on $N = 2\,499$ resamples is

$$(\hat{k}^*_{500,((N+1)0.05)}, \hat{k}^*_{500,((N+1)0.095)}) = (236, 256).$$

5.1.5 Multiple Changes

The detection of several change–points in multidimensional processes with unknown parameters can be done using the so called binary segmentation procedure. This method was proposed by Vostrikova [80] and enables to simultaneously detect the number and the location of the change–points. The method consists of first applying the likelihood ratio test from Section 5.1.1 for one change. If H_0 is rejected then we have the estimate of the time of the change $\hat{k}_n$. Next, we divide the sample in two subsamples $\{\mathbf{x}_i : 1 \leq i \leq \hat{k}_n\}$ and $\{\mathbf{x}_i : \hat{k}_n < i \leq n\}$ and test H_0 in each one of them. If we find a change in any of the sets we continue this segmentation procedure until we don't reject H_0 in any of the subsamples.

5.2 A comment on pricing

From the previous sections we have seen that standard change–point procedures and bootstrap methodology can be combined to come up with estimates for change–point structures in copula data. Of course, combining the copula with certain marginal distributions, this work can be extended to more general bivariate (or indeed multivariate) models on $\mathbb{R}^2$ (respectively $\mathbb{R}^d$). One of the reasons that a change–point analysis in dependence structures is important is because such changes often come from transitions from "normal" to "extreme" market conditions. Products priced correctly for instance for the former may be severely mispriced for the latter. In this section, and based on the example from Section 5.1.2, we will give a simple illustration of this. For further examples of applications of copula models as stress scenarios for financial positions in insurance or finance, see for instance [3].

	Before time change $\hat{\theta}_b = 0.8515$	After time change $\hat{\theta}_a = 0.4087$	Ignoring change–point $\hat{\theta}_0 = 0.6325$
$\hat{E}(\Psi_1)$	0.0315 (0.0001)	0.0556 (0.0001)	0.0453 (0.0001)
$\hat{E}(\Psi_1)/\hat{E}_{\hat{\theta}_0}(\Psi_1)$	0.70	1.23	1
$\hat{E}(\Psi_2)$	0.1245 (0.0005)	0.2646 (0.0006)	0.1972 (0.0006)
$\hat{E}(\Psi_2)/\hat{E}_{\hat{\theta}_0}(\Psi_2)$	0.63	1.34	1

Table 5.3: *Monte Carlo analysis for two positions on the (U, V) simulated data, considering the change–point or ignoring that change.*

We will consider bivariate risks (U, V) following three possible bivariate distributions. First we take the copula model from Section 5.1.2 with the change–point, and consider the two resulting models before and after the change. Recall that the estimated Gumbel parameters were $\hat{\theta}_b = 0.8515$ before and $\hat{\theta}_a = 0.4087$ after the change–point. The homogeneous model, ignoring the change–point, yielded an estimate $\hat{\theta}_0 = 0.6325$. As typical payout functions we take

$$\Psi_1(U, V) = (U + V - 1.5)_+ \quad \text{and} \quad \Psi_2(U, V) = (U + V)\mathbb{I}_{\{U>0.8, V>0.8\}}.$$

We can estimate the expected value of these two positions, $E(\Psi_1(U, V))$ and $E(\Psi_2(U, V))$, using simulation. Table 5.3 has the Monte Carlo estimates for these expected values under the three models. Each value was obtained from a simulated set of 1 000 000 pairs from the corresponding Gumbel copula. In brackets are the standard errors of the Monte Carlo estimations.

The estimated expected values vary considerably with the parameter of the distribution. From the table we can see that $\hat{E}(\Psi_1)$ after the time of the change is 1.23 of the value obtained if we ignore the change when performing the fitting. In the case of the second payout, the relative difference is even bigger, 1.34, between the estimated expected values considering and ignoring the change. This example illustrates the improvement that can be obtained when one recognises that there is a change–point in the data and one is able to take this change into account when modelling the data.

5.3 An example with insurance data

The Danish fire data is a set of $n = 2\,493$ trivariate observations consisting of losses to buildings, losses to contents and losses to profits. These data were analysed in a one–dimensional setting for instance in [27] and [61]. For a bivariate analysis, see [3]. The independence between the n observations seems to be an acceptable assumption. In our example we use the variables losses to contents and losses to profits and only the observations with strictly positive components. The resulting set has $n = 517$ bivariate observations. An analysis of this data set reveals that the Gumbel copula yields a good fit; see [3]. The parameter estimate of the copula is $\hat{\theta} = 0.5385$ and a Kolmogorov–Smirnov goodness of fit test gives that we do not reject the model with a p–value of 0.52.

Here we are interested in testing for the existence of changes in the parameter of the dependence structure. If we test for one change using (5.2) we obtain that $z_{517\,obs} = \max_{1\leq k<517}(-2\log(\Lambda_k)) = 3.4364$ which is attained at $k = 2$. The square root $z^{1/2}_{517\,obs} = 1.8537$ lies out of any rejection region for the usual levels. Actually, we could only reject H_0 for levels $(1-\alpha)$ smaller than 19.30%, using (5.4). In Figure 5.7 we plot the $(-2\log(\Lambda_k))$ for these data. The fact that the maximally selected likelihood ratio Z_{517} is attained at $k = 2$, as we saw in result (5.7), gives more confidence to not reject the no change hypothesis. This analysis strengthens the approach taken in Blum et al. [3] where functionals of the Danish data were priced using a homogeneous (i.e. no change) Gumbel model.

5.4 Change–point analysis of the FX returns copula

Based on the models from Chapter 3, in this section we test for the occurrence of change–points in the copula parameters of the FX filtered returns, deseasonalised in Chapter 2. Concretely we use the procedures from Section 5.1 to estimate change–points in the correlation parameter

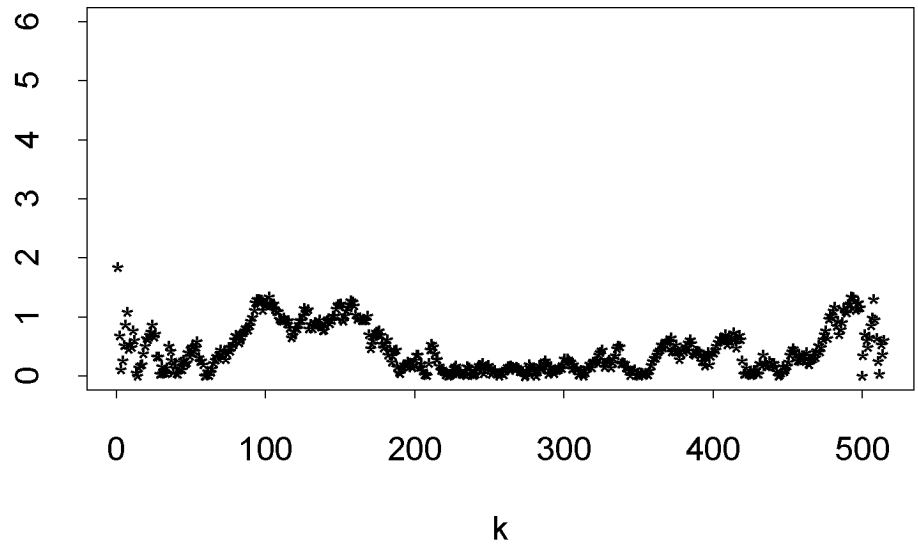

Figure 5.7: *Values of* $(-2\log(\Lambda_k))^{1/2}$, $1 \leq k < 517$, *for the Contents and Profits losses of the Danish fire data.*

of a t–copula fitted to the residuals of daily USD/DEM and USD/JPY returns. A similar analysis can be done at other frequencies. For the change–points found, we estimate the size of those changes and the corresponding time of occurrence. We also look at macro economic reasons possibly triggering the changes.

After filtering the univariate returns using the GARCH type models specified in Table 3.3 of Chapter 3, the residuals are assumed to be independent in time. Hence, we can use (5.2) and (5.5) for detecting possible change–points in the parameters of the multivariate contemporaneous conditional distribution and in particular in the copula.

We showed in Chapter 3 that for these residuals the t–copula yields the best fitting model for the dependence structure between the two series. For the copula fitting, we use the empirical distribution function to map the residuals into the unit square. Moreover, in a first step, we assume that the degrees of freedom of the copula are constant over time and hence we test for change–points in the correlation parameter. We evaluate Λ_k for $k = 1, 2, \ldots, n$ where $n = 3\,259$; see (5.1). The values obtained are displayed in the top panel of Figure 5.8. The test statistic (5.2) takes the value $z^{1/2}_{n\,obs} = 13.26$ and by (5.5) we have that $P(Z^{1/2}_n > 13.26) \approx 0$. The

null hypothesis of no change–point is to be rejected and the estimated time of the change is $\hat{k}_n$ = 8 November 1989; corresponding to the fall of the Berlin wall. In the next step, the sample is divided in two sub–samples, one up to November, 8 1989 and another from the estimated time of change onwards. For each sub–sample Λ_k is computed as well as $Z_n^{1/2}$. The middle panel of Figure 5.8 plots these estimates and Table 5.4 has the values for $z_{n\,obs}^{1/2}$ and all the information about the testing procedure. As the obtained p–values are close to zero we reject the null hypothesis of no change for each sub–sample and estimate two more times of change, December, 29 1986 and June, 18 1997. The later date corresponds to the beginning of the Asia crisis starting with the violent devaluation of the Thai Baht. Each sub–sample is again divided in two and the procedure is repeated yielding the estimates in the bottom panel of Figure 5.8.

$z_{n\,obs}^{1/2}$	n	$P\left(Z_n^{1/2} > z_{n\,obs}^{1/2}\right)$	$H_0(0.95)$	Time of change
13.26	3 259	0	reject	8 Nov. 1989
5.96	923	0.0000004	”	29 Dec. 1986
5.31	2 336	0.0000143	”	18 June 1997
2.99	176	0.0689621	not rej.	(23 June 1986)
3.10	747	0.0709747	”	(31 July 1989)
5.86	1 985	0.0000007	reject	23 Oct. 1990
2.36	351	0.3380491	not rej.	(8 Sep. 1998)
2.78	1 736	0.1873493	”	(21 Oct. 1996)
2.86	249	0.1061709	”	(21 Mars 1990)

Table 5.4: *Change–point analysis for USD/DEM and USD/JPY spot rate residuals.*

For the results of the analysis showed in the bottom panel of Figure 5.8 only for the maximum attained at October, 23 1990 the null hypothesis is rejected at a 95% level. So we still have to split this sub–sample further. The first from November, 8 1989 until October, 23 1990 and the second from this date up to June, 18 1997. The $z_{n\,obs}^{1/2}$ obtained in these cases are low, see Table 5.4, and we do not reject the null hypothesis of no change in both cases. In Table 5.4 we give the time where the test statistic is attained for each sub–sample. If the null hypothesis is not rejected the referred date is in parentheses as it is not considered a

change–point. In summary we found four change–points: December, 29 1986, November, 8 1989, October, 23 1990 and June, 18 1997. For the five periods between the times of change we estimated for the copula correlation ($s\hat{e}.$): $\hat{\rho}_1 = 0.6513$ (0.0384), $\hat{\rho}_2 = 0.8312$ (0.0113), $\hat{\rho}_3 = 0.3099$ (0.0608), $\hat{\rho}_4 = 0.5752$ (0.0149) and $\hat{\rho}_5 = 0.3505$ (0.0460). To visualise these results we redisplay in Figure 5.9 the time–varying correlation estimated in Chapter 4, first plotted in the bottom panel of Figure 4.4. On these estimates we super–impose the estimated change–point cross–correlation for the five periods between the times of change. The two estimated paths do not disagree. The change–point analysis detects big changes sooner but it is less sensitive to the small ones. Actually, no change–points were detected from October, 23 1990 until June, 18 1997 which seems perhaps quite a long period for the correlation to be constant. It may be interesting to note that the former date (October, 23 1990) corresponds to the burst in the Japanese asset price bubble. On October, 18 1990, the USD/JPY ended a fall from about 158 to 125.

Comments

In this chapter, we have looked at model changes within a parametric copula set up. Besides explaining how the classical literature can be adapted to detect change–points in copula (hence dependence) parameters, we have also discussed the construction of bootstrap confidence intervals. We have illustrated the methodology introduced on simulated and real data. Namely, we tested for change–points in the iid Danish fire data and in the residuals of the time dependent deseasonalised FX log–returns on USD/DEM and USD/JPY. The tests did not reveal change–points for the insurance data but found four breaks in the correlation for the FX data set. Moreover, economic events that may have triggered the estimated changes were identified. It would still be interesting in future work, to analyse the existence of change–points in the parameters of the equations that define the copula dynamics in the models with time–varying dependence parameters from Chapter 4.

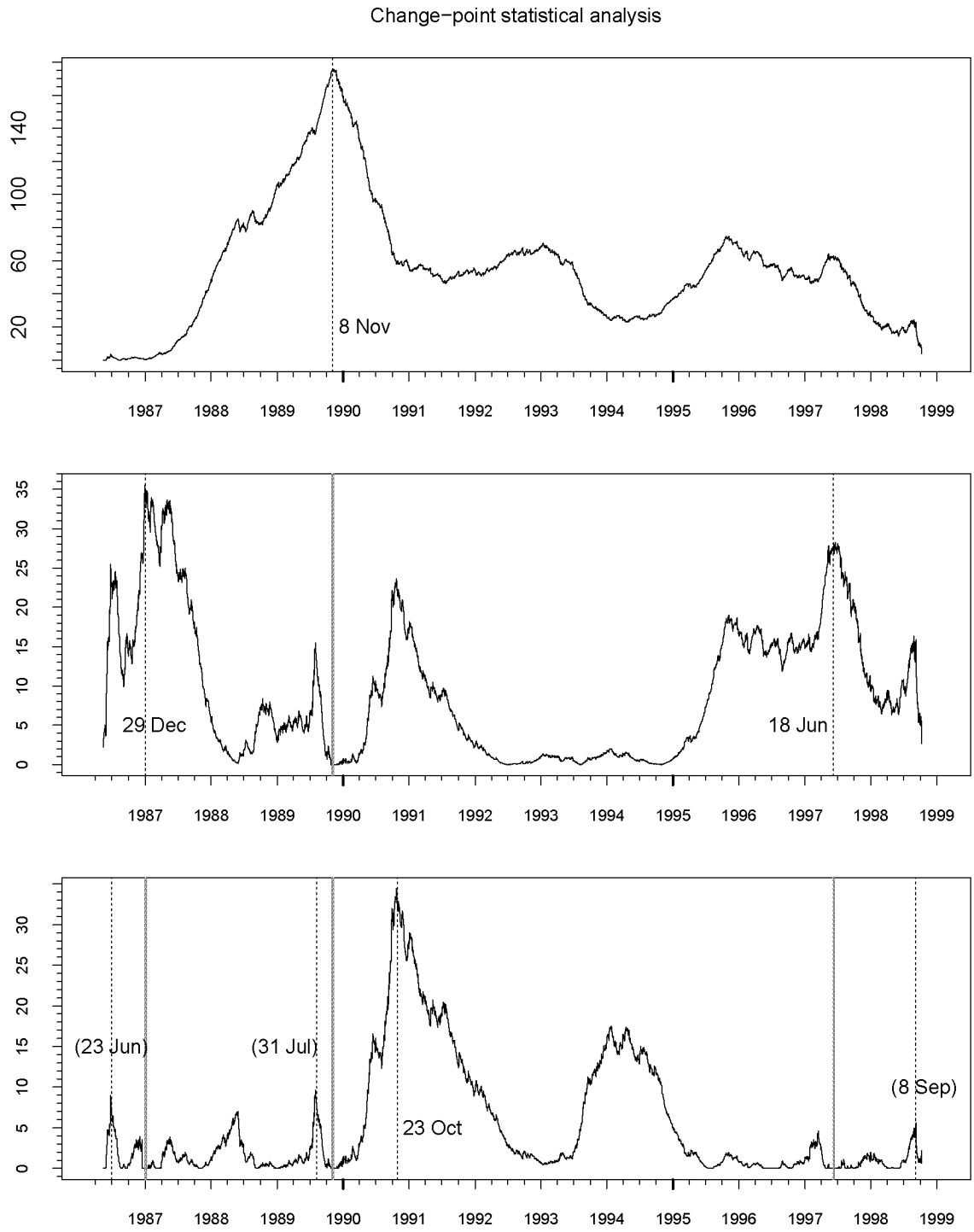

Figure 5.8: *Change–point analysis of daily returns on the FX rates USD/DEM and USD/JPY spot rates. The three panels display three steps of the change–point analysis. Each panel plots the likelihood ratio values* $-2\log(\Lambda_k)$ *for* $k = 1, 2, \ldots, n$. *In each sub–sample its maximum, the test statistic* Z_n, *gives the time of the change in case the no change null hypothesis is rejected. If the null hypothesis is not rejected the moment where* Z_n *is achieved is put in parentheses.*

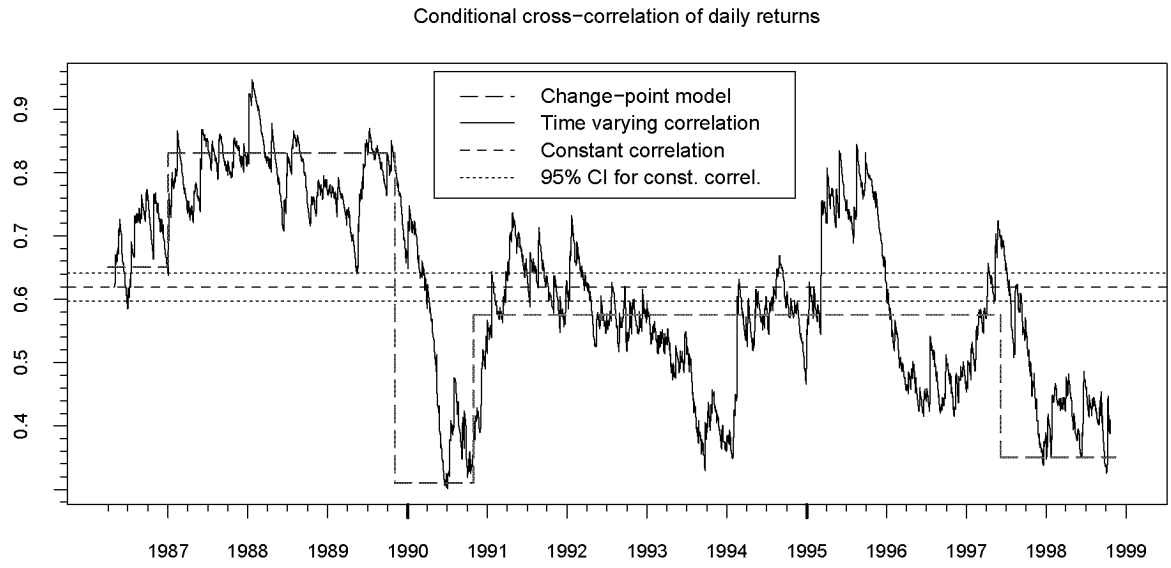

Figure 5.9: *Estimated t–copula correlation paths of daily returns on the FX USD/DEM and USD/JPY spot rates. The long–dashed line is the estimated correlation by the change–point tests. This is super–imposed on the estimated correlation using the time–varying copula model from Chapter 4. The short–dotted line is the time–invariant correlation estimate after marginal GARCH filtering. The change–points model seems to react quicker to important economic events than the time–varying copula model but ignores smaller changes, at least at the daily frequency.*

Summary and conclusion

Within finance and insurance, a considerable amount of effort is more recently put on the modelling of multivariate (typically dependent) data and the pricing of instruments based on such data. Copula–based models are widely accepted as providing an interesting approach for the construction of such models looking at dependence beyond linear correlation. Our goal was to bring together and introduce the tools suitable for statistical copula analysis of data from finance and insurance.

In the first chapter we present copula related definitions and properties together with relevant examples. Two inference methods for copula–based models are described: the IFM and the pseudo log–likelihood methods. In each case standard errors for the model parameter estimates can be computed. A goodness–of–fit test is worked out. These notions are illustrated for weekly log–returns of the FX rates USD/DEM and USD/JPY. From several families a mixture of Gumbel with survival Gumbel copulae and the t–copula provides the best models for these data. We exemplify how these models can be used to estimate risk measures like VaR and expected shortfall for a hypothetic portfolio. The computation of bootstrap confidence intervals for these measures is also presented and a backtesting exercise is performed. Less computer intensive, the construction of jackknife confidence intervals for the tail–dependence point estimates is

explained. In summary, Chapter 1 presents and illustrates the main inference tools for copula–based models.

In Chapter 2 we are more ambitious and start tackling the modelling of high–frequency financial data. Our goal is twofold. First exemplify the statistical models introduced and second (more importantly) study the stylised facts of tick–by–tick bivariate FX data. A first challenge is the complicated weekly and intra–day seasonal pattern of the data. A deseasonalisation method is applied to the five minutes FX log–returns on USD/DEM and USD/JPY spot rates. We hence obtain deseasonalised log–returns for six different time horizons from one hour up to one day. Ignoring at first the serial or time dependence, we use the pseudo log–likelihood method to fit copula–based models to these data. The t–copula provides the best model at all frequencies and based on it we obtain tail–dependence estimates revealing the existence of asymptotic tail–dependence in the data. Moreover we present a test for the elliptical symmetry of the data which is not rejected for the lowest frequencies. Using the notion of regularly varying tail distribution from multivariate extreme value theory, we can estimate the spectral density for each frequency. The excesses beyond a given bivariate threshold are modelled for the first and third quadrants and the survival Clayton and Clayton copulae, respectively, are found to be good models. This agrees with the corresponding theoretical asymptotic theory.

The next problem to be taken care of is time dependence. We use univariate GARCH type models in order to model the serial dependence. In Chapter 3 several time–invariant copula models are fitted to the filtered returns. At this residual level, the t–copula again turns out to be a potentially good model followed by a mixture of a Gumbel with a survival Gumbel for all frequencies except for the daily returns where the two models perform almost equally well, with slight advantage for the mixture one. Ellipticity of the residuals is rejected only for one and two hour frequencies when the margins are t transformed.

Chapter 4 starts by investigating the time homogeneity of the linear correlation between the USD/DEM and USD/JPY deseasonalised returns for

the six frequencies using bivariate matrix–diagonal GARCH models. The fitting results point very clearly to dynamics in the dependence structure. As a model, the combination of univariate GARCH type models with a copula family with time–varying parameters is analysed. We call the resulting model a time–varying copula model to distinguish it from the time–invariant copula models from Chapter 3. Using the t–copula as a model for the dependence structure we propose time dynamics for the dependence parameters (degrees of freedom and correlation coefficient). Once fitted to the data, the estimated correlation path reveals an interesting structure. Economic events that might have triggered some of the features estimated were identified.

A further tool useful for dynamic copula modelling consists of change–point detection. Estimating change–points in the copula parameters and then estimating these parameters separately between the times of change we obtain piecewise constant copula parameter estimates. In Chapter 5 we test the existence of change–points in the copula parameters. Applying the method to a real insurance data set no change–points were detected. On the other hand, testing for change–points in the copula of the daily filtered deseasonalised log–returns on the FX data set, four changes were detected. For three of these it was possible to find economic events which may have triggered the changes.

Clearly more general copula models can be handled. Moreover, in models with time–varying dependence parameters, the existence of change–points can also be tested in the parameters of the equations that define the copula dynamics. The applicability of non–parametric test statistics for change–points could also be studied. We will definitely return to some of these questions in future work.

The software used for the computations were S–Plus, the package S+FinMetrics and C^{++}. A considerable number of algorithms was tested and implemented for the simulation and estimation of copula families, goodness–of–fit testing and detection of change–points in copulae. As stated in the Introduction, we chose not to include these programs in the thesis; details are available upon request.

Bibliography

[1] Anderson, T. W., and Darling, D. A. A test of goodness of fit. *J. Amer. Statist. Assoc. 49* (1954), 765–769.

[2] Andreou, E., and Ghysels, E. Rolling–sample volatility estimators: some new theoretical, simulation, and empirical results. *J. Bus. Econom. Statist. 20*, 3 (2002), 363–376.

[3] Blum, P., Dias, A., and Embrechts, P. The art of dependence modelling: the latest advances in correlation analysis. In: *Alternative Risk Strategies* (Ed. Morton Lane), Risk Books, London (2002), 339–356.

[4] Bollerslev, T., Chou, R. Y., and Kroner, K. ARCH modeling in finance. *J. Econometrics 52* (1992), 5–59.

[5] Bollerslev, T., Engle, R. F., and Nelson, D. B. Arch models. In *Handbook of econometrics, Vol. IV*, vol. 2 of *Handbooks in Econom.* North–Holland, Amsterdam, 1994, pp. 2959–3038.

[6] Borio, C., Furline, C., and Lowe, P. Procyclicality of the financial system and financial stability: issues and policy options. Working paper, BIS, Basel, 2001.

[7] Boyer, B. H., Gibson, M. S., and Loretan, M. Pitfalls in tests for changes in correlations. *Board of Governors of the Federal Reserve System, International Finance Discussion Papers 597* (1999).

[8] Breymann, W., Dias, A., and Embrechts, P. Dependence structures for multivariate high–frequency data in finance. *Quant. Finance 3* (2003), 1–14.

[9] Brockwell, P. J., and Davis, R. A. *Time Series: Theory and Methods*, second ed. Springer Series in Statistics. Springer–Verlag, New York, 1991.

[10] Cappiello, L., Engle, R., and Sheppard, K. Asymetric dynamics in the correlations of global equity and bond returns. Working paper 204, European Central Bank, 2003.

[11] Charpentier, A. Tail dependence for Archimedian copulas. Working paper, CREST, Laboratoire de Finance et Assurance, 2003.

[12] Chen, X., and Fan, Y. Estimation of copula–based semiparametric time series models. Working paper, New York University and Vanderbilt University, 2002.

[13] Clayton, D., and Cuzick, J. Multivariate generalizations of the proportional hazards model. *J. Roy. Statist. Soc. Ser. A 148*, 2 (1985), 82–117.

[14] Csörgő, M., and Horváth, L. *Limit Theorems in Change–Point Analysis.* Wiley, Chichester, 1997.

[15] Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., and Pictet, O. V. *An Introduction to High–Frequency Finance.* Academic Press, San Diego, CA, 2001.

[16] Darsow, W. F., Nguyen, B., and Olsen, E. T. Copulas and Markov processes. *Illinois J. Math. 36*, 4 (1992), 600–642.

[17] Davison, A. C., and Hinkley, D. V. *Bootstrap Methods and Their Application.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.

[18] Dias, A., and Embrechts, P. Change–point analysis for dependence structures in finance and insurance. In: *Novos Rumos em Estatística* (Ed. C. Carvalho, F. Brilhante and F. Rosado), Sociedade

Portuguesa de Estatística, Lisbon (2002), 69–86. Also to appear in: *New Risk Measures in Investment and Regulation* (Ed. G. Szego), John Wiley and Sons, New York (2003).

[19] Dias, A., and Embrechts, P. Dynamic copula models for multivariate high–frequency data in finance. Working paper, ETH Zurich, Departement of Mathematics, 2003.

[20] Ding, Z., Granger, C. W. J., and Engle, R. F. A long memory property of stock market returns and a new model. *J. Empirical Finance 1* (1993), 83–106.

[21] Eberlein, E., and Raible, S. Some analytic facts on the generalized hyperbolic model. In *European Congress of Mathematics, Vol. II (Barcelona, 2000)*, vol. 202 of *Progr. Math.* Birkhäuser, Basel, 2001, pp. 367–378.

[22] Efron, B. *The Jackknife, the Bootstrap and other Resampling Plans*, vol. 38 of *CBMS–NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1982.

[23] Efron, B., and Tibshirani, R. J. *An Introduction to the Bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

[24] Embrechts, P., Ed. *Extremes and Integrated Risk Management*. RISK Waters Group, London, 2000.

[25] Embrechts, P. Extremes in economics and the economics of extremes. In: *Extreme Values in Finance, Telecommunications, and the Environment* (Ed. B. Finkenstadt and H. Rootzen), Chapman & Hall/CRC, London (2003), 169–183.

[26] Embrechts, P., Frey, R., and McNeil, A. J. Stochastic Methods for Quantitative Risk Management. Book manuscript (2003).

[27] Embrechts, P., Klüppelberg, C., and Mikosch, T. *Modelling Extremal Events for Insurance and Finance*. Springer–Verlag, Berlin, 1997.

[28] Embrechts, P., Lindskog, F., and McNeil, A. J. Modelling dependence with copulas and applications to risk management. In: *Handbook of Heavy Tailed Distributions in Finance* (Ed. S.T. Rachev), Elsevier/North–Holland (2003), 329–384.

[29] Embrechts, P., McNeil, A. J., and Straumann, D. Correlation and dependence in risk management: Properties and pitfalls. In: *Risk Management: Value at Risk and Beyond* (Ed. M. Dempster), Cambridge University Press, Cambridge (2002), 176–223.

[30] Engle, R., and Sheppard, K. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working paper 2001–15, UCSD, 2001.

[31] Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 4 (1982), 987–1007.

[32] Engle, R. F., and Kroner, K. F. Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 1 (1995), 122–150.

[33] Fortin, I., and Kuzmics, C. Tail dependence in stock return–pairs: Towards testing ellipticity. Working paper, Institute For Advanced Studies, Vienna and Faculty of Economics and Politics, University of Cambridge, 2002.

[34] Foster, D. P., and Nelson, D. B. Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*, 1 (1996), 139–174.

[35] Frees, E. W., and Valdez, E. Understanding relationships using copulas. *N. Am. Actuar. J. 2*, 1 (1998), 1–25.

[36] Genest, C. Frank's family of bivariate distributions. *Biometrika 74*, 3 (1987), 549–555.

[37] Genest, C., Ghoudi, K., and Rivest, L.-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*, 3 (1995), 543–552.

[38] Genest, C., and Rivest, L.-P. Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc. 88*, 423 (1993), 1034–1043.

[39] Genest, C., and Werker, B. J. M. Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In: *Distributions with Given Marginals and Statistical Modelling* (Ed. C. M. Cuadras, J. Fortiana and J. A. Rodriguez–Lallena), Kluwer Academic Publishers, Dordrecht (2002), 103–112.

[40] Gombay, E., and Horváth, L. On the rate of approximations for maximum likelihood test in change–point models. *J. Multivariate Anal. 56*, 1 (1996), 120–152.

[41] Gombay, E., and Horváth, L. Change–points and bootstrap. *Environmetrics 10* (1999), 725–736.

[42] Hall, P. *The Bootstrap and Edgeworth Expansion.* Springer–Verlag, New York, 1992.

[43] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York, 1986.

[44] Hartmann, P., Straetmans, S., and de Vries, C. G. Asset market linkages in crisis periods. Working paper, no. 71, European Central Bank, Frankfurt, 2001.

[45] Hauksson, H. A., Dacorogna, M. M., Domenig, T., Müller, U. A., and Samorodnitsky, G. Multivariate extremes, aggregation and risk estimation. *Quant. Finance 1* (2001), 79–95.

[46] Hougaard, P. Fitting a multivariate failure time distribution. *IEEE Trans. Reliab. 38* (1989), 444–448.

[47] Jarque, C. M., and Bera, A. K. A test for normality of observations and regression residuals. *Internat. Statist. Rev. 55*, 2 (1987), 163–172.

[48] Joe, H. *Multivariate Models and Dependence Concepts.* Chapman & Hall, London, 1997.

[49] Juri, and Wüthrich. Copula convergence theorems for tail events. *Insurance Math. Econom. 30* (2002), 405–420.

[50] Kiefer, N., and Vogelsang, T. A new asymptotic theory for heteroskedasticity–autocorrelation robust tests. Working paper, Cornell University, 2002.

[51] Klaassen, C. A. J., and Wellner, J. A. Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli 3*, 1 (1997), 55–77.

[52] Klugman, S. A., and Parsa, R. Fitting bivariate loss distributions with copulas. *Insurance Math. Econom. 24* (1999), 139–148.

[53] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics 54* (1992), 159–178.

[54] Lehmann, E. L. *Theory of Point Estimation.* Wadsworth & Brooks, Pacific Grove, California, 1991.

[55] Lehmann, E. L., and Casella, G. *Theory of Point Estimation*, second ed. Springer Texts in Statistics. Springer–Verlag, New York, 1998.

[56] Lindskog, F., McNeil, A. J., and Schmock, U. Kendall's tau for elliptical distributions. In: *Credit Risk–Measurement, Evaluation and Management* (Ed. G. Bol, G. Nakhaeizadeh, S. T. Rachev, T. Ridder and K.–H. Vollmer), Physica–Verlag, Heidelberg (2003), 149–156.

[57] Longin, F., and Solnik, B. Extreme correlation of international equity markets. *J. Finance LVI*, 2 (2001), 649–676.

[58] Loretan, M., and Phillips, P. C. B. Testing the covariance stationarity of heavy–tailed time series: An overview of the theory

with applications to several financial data sets. *J. Empirical Finance 2* (1994), 211–248.

[59] Manzotti, A., Pérez, F. J., and Quiroz, A. J. A statistic for testing the null hypothesis of elliptical symmetry. *J. Multivariate Anal. 81*, 2 (2002), 274–285.

[60] McLeish, D. L., and Small, C. G. *The Theory and Applications of Statistical Inference Functions*, vol. 44 of *Lecture Notes in Statistics*. Springer–Verlag, New York, 1988.

[61] McNeil, A. J. Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin 27*, 1 (1997), 117–137.

[62] Müller, U. A., Dacorogna, M. M., and Pictet, O. V. Heavy tails in high–frequency financial data. In: *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions* (Ed. R. J. Adler, R. E. Feldman and M. S. Taqqu), Birkhäuser, Boston (1998), 55–77.

[63] Nelsen, R. B. *An Introduction to Copulas*, vol. 139 of *Lecture Notes in Statistics*. Springer–Verlag, New York, 1999.

[64] Oakes, D. A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B 44*, 3 (1982), 414–422.

[65] Oakes, D. Semiparametric inference in a model for association in bivariate survival data. *Biometrika 73*, 2 (1986), 353–361.

[66] Oakes, D. Multivariate survival distributions. *J. Nonparametr. Statist. 3*, 3–4 (1994), 343–354.

[67] Patton, A. J. Modelling time–varying exchange rate dependence using the conditional copula. Working paper, UCSD, 2002.

[68] Pham-Gia, T., and Hung, T. L. The mean and median absolute deviations. *Mathematical and Computer Modelling 34* (2001), 921–936.

[69] Platen, E. Arbitrage in continuous complete markets. *Adv. in Appl. Probab. 34*, 3 (2002), 540–558.

[70] Resnick, S. *Extreme Values, Regular Variation, and Point Processes.* Springer–Verlag, New York, 1987.

[71] Resnick, S. On the foundations of multivariate heavy tail analysis. Cornell Reports and Papers, 2002. Available on–line from `http://www.orie.cornell.edu/~sid`.

[72] Rockinger, M., and Jondeau, E. Conditional dependency of financial series: An application of copulas. Working paper, HEC–School of Management, Departement of Finance, 2001.

[73] Rosenblatt, M. Remarks on a multivariate transformation. *Ann. Math. Statistics 23* (1952), 470–472.

[74] Schafer, J. L. *Analysis of Incomplete Multivariate Data*, vol. 72 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, London, 1997.

[75] Schweizer, B., and Wolff, E. F. On nonparametric measures of dependence for random variables. *Ann. Statist. 9*, 4 (1981), 879–885.

[76] Shao, J., and Tu, D. S. *The Jackknife and Bootstrap.* Springer Series in Statistics. Springer–Verlag, New York, 1995.

[77] Shephard, N. Statistical aspects of ARCH and stochastic volatility. In: *Time Series Models* (Ed. D. R. Cox and D. V. Hinkley), Chapman and Hall Ltd, London (1996), 1–67.

[78] Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8* (1959), 229–231.

[79] Stărică, C. Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance 6* (1999), 515–553.

[80] Vostrikova, L. J. Detecting "disorder" in multidimensional random processes. *Soviet Mathematics Doklady 24*, 1 (1981), 55–59.

[81] Wang, S. Aggregation of correlated risk portfolios: Models and algorithms. Preprint, Casualty Actuarial Society (CAS), 1997.

[82] Zivot, E., and Wang, J. *Modeling Financial Time Series with S–Plus.* Springer–Verlag, New York, 2003.

List of Figures

List of Tables

Curriculum Vitae

Personal Data

Name:	Alexandra da Costa Dias
Date of Birth:	31.07.1967
Nationality:	Portuguese

Education

2000 – date:	Ph.D. student in Mathematics at ETH Zurich; Supervisor: Prof. Dr. Paul Embrechts
1996 – 1998:	Postgraduate student of Actuarial Mathematics and Financial Risk Management, ISEG, Technical University of Lisbon; M.Sc. awarded in February 1998
1988 – 1994:	Undergraduate student of Mathematics, Faculty of Sciences and Technology, New University of Lisbon; Licentiate degree awarded in July 1994

Employment

1998 – date:	Teaching Assistant in the Department of Mathematics of Faculty of Sciences and Technology, New University of Lisbon
1995 – 1998:	Trainee Teaching Assistant in the Department of Mathematics of Faculty of Sciences and Technology, New University of Lisbon
1994 – 1995:	Teaching Monitor in the Department of Mathematics of Faculty of Sciences and Technology, New University of Lisbon
1993 – 1994:	Trainee actuary with Fungest, Sociedade Gestora de Fundos de Pensões SA, Lisbon