Multivariate Extremes, Aggregation and Risk Estimation *

Höskuldur Ari Hauksson

Michel Dacorogna

Thomas Domenig

Ulrich Müller

Gennady Samorodnitsky[†]

May 12, 2000

*Olsen & Associates, Research Institute for Applied Economics, Seefeldstrasse 233,
CH-8008 Zürich, Switzerland. e-mail: `research@olsen.ch`

[†]School of Operations Research and Industrial Engineering and Department
of Statistical Science, Cornell University, Ithaca, NY 14853, USA. e-mail:
`gennady@orie.cornell.edu`

Contents

1	Introduction	1
2	Extreme Value Theory	2
2.1	Multivariate Extreme Value Theory	3
2.2	Tails Under Aggregation	6
3	Exploring the Empirical Tails	10
3.1	Spatial Dependence	10
3.2	Serial Dependence	18
4	Risk Management	24
5	Conclusion	30

Abstract

We briefly introduce some basic facts about multivariate extreme value theory and present some new results regarding finite aggregates and multivariate extreme value distributions. Based on our results high frequency data can considerably improve quality of estimates of extreme movements in financial markets.

Secondly we present an empirical exploration of what the tails really look like for four foreign exchange rates sampled at varying frequencies. Both temporal and spatial dependence is considered. In particular we estimate the spectral measure, which along with the tail index, completely determines the extreme value distribution.

Lastly we apply our results to the problem of portfolio optimisation or risk minimization. We analyze how the expected shortfall and VaR scale with time horizon and find that this scaling is not by a factor of square root of time as is frequently used, but by a different power of time. We show that the accuracy of risk estimation can be drastically improved by using hourly or bihourly data.

1 Introduction

It is generally accepted among practitioners that financial assets are more dependent when the market is in an excited state than when it is not. What is meant by "more dependent" and "excited state" is at this point rather vague. Our goal is to clarify this statement by exploring what historical data really looks like. This includes analyzing serial and spatial (between assets) dependence as well as behavior under aggregation. Our results have direct implications for various aspects of risk management.

Traditionally the benchmark measure of risk associated with some cash flow has been its variance. Current best practices, however, use quantile based methods to estimate this risk. Of those Value-at-Risk (VaR) is undoubtedly the most popular risk measure to date. The VaR does not behave well with respect to diversification [2], [3], and to overcome this shortcoming another risk measure, the expected shortfall (ES) (alias conditional VaR, CVaR, mean excess), has been proposed.

Whether VaR, ES or some other measure of risk will be dominant in the near future, it is clear that the emphasis in risk management has been shifted from estimating the risk based on the full return distribution to focusing exclusively on the highest return quantiles. Given this fact, it is clear that studying the dependence structure, both spatial and temporal, for the high quantile regime is crucial to further the understanding of financial risk.

We attempt to do both here by estimating spectral measures, exceedence probabilities and some other features of the tails of the return distribution.

Extreme events are by definition rare. Estimating probabilities at high quantiles is often a difficult task which requires large data sets and the use of extreme value theory. The tails of financial processes have been studied at length in the univariate setting (see e.g. [1]). In the multivariate setting the problem of lack of data becomes even larger. Instead of using data collected over longer time intervals, we propose to sample the price process at a higher frequency. Few multivariate tail analysis studies are available, of which we would like to mention [6], [22], [28], [29].

The paper is organized as follows: We first give a brief introduction to multivariate extreme value theory. Here we try to limit ourselves to those concepts that will be needed in the paper. We also present some results on the behavior of extremes under finite aggregation. These results are important for our application as they provide the necessary connection between extremes of returns sampled at various frequencies.

Next section explores the dependence structures between the assets in our dataset. Here we consider both spatial and temporal dependence of the extreme events. In particular, we estimate the spatial and temporal spectral measure and the exceedence probabilities, as well as estimating how the tails are different from those of elliptical distributions and regularly varying distributions. Furthermore, our study covers several return time interval sizes so it provides a rich picture of what the tails of foreign exchange return processes look like. From our empirical exploration we can conclude that the assumptions of the theory in the previous section are indeed satisfied.

We show that the risk as measured by VaR and ES scales in time like the time interval to a certain power. This power is, contrary to current practices, not 0.5 but a slightly smaller number. As a result of this scaling law, one can estimate the risk with high frequency data, where more data provide for better estimates.

2 Extreme Value Theory

Using methods of extreme value theory has been popular in the finance literature in recent years. This has gone hand in hand with the increased interest in quantile based methods for risk estimation. The bulk of this research has focused on the univariate case; we have mentioned above some examples of multivariate analysis. We first give a brief introduction to multivariate EVT

in the iid setting. We then discuss dependent sequences and extreme value theory for finite sums of random processes. For a thorough discussion of univariate extreme value theory see [14], [21] or [26]. For a discussion of the multivariate case see [10], [12] or [18]. A discussion of a broad range of different dependence functions in two dimensions is given in [20].

2.1 Multivariate Extreme Value Theory

Consider the first n drawings from an iid multivariate process (X_i) with a multivariate cumulative distribution function F . Let $X_{(1)}, \dots, X_{(n)}$ denote the ranking in descending order of the first n drawings, where we rank in each coordinate separately. Let $X_i(m)$ and $X_{(i)}(m)$ denote the m -th coordinate of X_i and $X_{(i)}$ respectively, $m = 1, \dots, d$. Suppose that there exists a distribution function G with non-degenerate marginals and sequences of constant vectors $a_n > 0$ and b_n such that

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[\frac{X_{(1)}(1) - b_n(1)}{a_n(1)} \leq x_1, \dots, \frac{X_{(1)}(d) - b_n(d)}{a_n(d)} \leq x_d \right] \\ = \lim_{n \rightarrow \infty} F^n(a_n \cdot x + b_n) \\ = G(x) \end{aligned}$$

for all $x = (x_1, \dots, x_d)$. We then say that G is an *extreme value distribution* (EVD) and that F lies in the domain of attraction of G . Note that the limiting function is unique up to a scale and location transformation. This is clear, as the sequences a_n and b_n can be chosen so as to change the scale and location of G . We will say that two extreme value distributions G_1 and G_2 are of the same *type* if one is a scaled and translated copy of the other, i.e. if there exist constant vectors $a > 0$ and b such that $G_1(x) = G_2(a \cdot x + b)$.

It turns out that the only possible limiting distribution functions are the so called *max-stable* distribution functions. A distribution function is said to be max-stable if there exist real vector valued functions $a(s) > 0$ and $b(s)$ defined for all $s > 0$ such that

$$G^s(a(s) \cdot x + b(s)) = G(x)$$

for all $s > 0$ and d -dimensional vectors x .

In one dimension we have the celebrated Fisher-Tippett Theorem which states that if the extreme value distribution exists then it is either a Fréchet, Weibull or a Gumbel distribution, all of which are determined by a single

parameter. These distributions can all be written as the *generalized extreme value distribution* which is given by

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

(in the range $1 + \xi x > 0$ if $\xi \neq 0$). If $\xi > 0$ (the Fréchet extreme value distribution) then $1/\xi = \alpha$ is the *tail index*. Note that some other authors refer to ξ as the tail index and α as the shape index. The generalized extreme value distribution with $\xi < 0$ is called the Weibull distribution, and in the case $\xi = 0$ it is referred to as the Gumbel distribution.

For an EVD G , the set of all distributions which have G as an EVD is called the *maximum domain of attraction*, denoted by $MDA(G)$. In the univariate iid setting the maximum domains of attraction of the different EVD can be completely characterized.

It is generally accepted that the distribution of financial returns belong to the (maximum) domain of attraction of the Fréchet extreme value distributions, see e.g. [25]. These distributions have heavy tails and not all moments exist. Examples of distributions in the domain of attraction of the Fréchet distribution are the stable distributions and Student-t distributions. The Gumbel distribution has a thin tail and all moments exist. The normal distribution lies in the domain of attraction of the Gumbel distribution. The Weibull distribution has a right tail with a finite endpoint, and so do the distributions in its domain of attraction. An example of a distribution in the domain of attraction of the Weibull extreme value distribution is the univariate uniform distribution on the interval $[0, 1]$.

In the multivariate setting the problem of classifying possible EVDs is more complicated. It turns out that a given d -dimensional extreme value distribution is completely determined by its univariate marginal extreme value distributions and a finite measure (often called spectral, or tail, measure) that lives on the d -dimensional unit sphere, which describes the dependence between the variables [12]. This measure lives in a $d - 1$ dimensions, unlike the *copula* (see [18]), which lives in d dimensions. In particular, in two dimensions the unit sphere is commonly identified with the interval $[-\pi, \pi]$, and a finite measure on the unit sphere with its cumulative distribution function. Hence, the dependence structure in two dimensions is completely determined by this one dimensional dependence function. We will say a bit more about spectral measures below.

In general, using the EVD to make inference of properties of the original distribution F involves making some assumptions. Each extreme value distribution G can have a large domain of attraction, i.e. a large set of distribution functions F which all give rise to the same extreme value distribution G . These distributions may be quite different and even their tails, which are of prime interest to us, are only similar, and not quite the same. A common assumption in the univariate case is that the right hand tail of F is asymptotically given by $1 + \log(G)$. This assumption eliminates slowly varying functions (see below). Sometimes one is even willing to assume that the two tails coincide from some point on.

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

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

where $L(x)$ is a slowly varying function, i.e. $L(cx)/L(x) \rightarrow 1$ for all $c > 0$ as $x \rightarrow \infty$. In higher dimensions a distribution is *regularly varying* if there exists a constant $\alpha > 0$ and a random vector Θ with values on $\mathbf{S}^{d-1}$, the unit sphere in $\mathbf{R}^d$, such that the following limit exists for all $x > 0$

$$\frac{P[\|X\| > tx, X/\|X\| \in \cdot]}{P[\|X\| > t]} \xrightarrow{v}_{t \rightarrow \infty} x^{-\alpha} P_{\Theta}(\cdot) \quad (1)$$

where $\rightarrow^v$ denotes *vague convergence* on $\mathbf{S}^{d-1}$ and P_{Θ} is the distribution of Θ . A sequence of probability measures (μ_n) is said to converge vaguely to a probability measure μ if for all relatively compact sets A such that $\mu(\partial A) = 0$ we have $\mu_n(A) \rightarrow \mu(A)$. Intuitively, what regular variation means is that asymptotically the distribution can in polar coordinates be represented by a product measure, which is the product of the *spectral measure* P_{Θ} and a radial measure which has a power decay. For other definitions of regularly varying functions in the multivariate setting see [30].

If X is a regularly varying random variable, then it is in the domain of attraction of an EVD with Fréchet marginals, and it is completely determined by the spectral measure and the index α . Note that EVDs, as we have defined them here, have their spectral measure in the positive orthant, i.e. $[0, \infty]^d$. This is because we are one deals with the maxima of many random variables. If one replaces sum of the coordinate maxima by the corresponding minima then one can get an EVD in other orthants. These may, in genrela, have a different tail index α and a different spectral measure. It is also important to mention that an EVD may have a different tail index in different directions.

So far we have only considered iid processes. Clearly, one can not assume that financial processes are iid. In the case of a serially dependent process (X_n) there are two questions that need to be addressed: Does there exist an EVD for its maximum? And if so, does it have the same EVD type as obtained by treating the process as if it were an iid process? The process which is iid, but with the same marginal cumulative distribution functions is called the *associated process*, denoted by $(\tilde{X}_n)$.

The answer to the first question is frequently affirmative for stationary sequences with serial dependence that decays fast enough. For results in one dimension see Leadbetter *et al.* [21] and their condition D .

The answer to the second question tends to depend on the amount of clustering of extremes. In one dimension, under certain conditions, see condition D' of Leadbetter *et al.* [21], one can show that the limiting EVD of (X_n) and its associated process are identical. In many other cases one is a scaled and translated copy of the other, (when one uses the same sequences of scale and shift constants (a_n) and (b_n)), i.e. they have the same type, see Embrechts *et al.* [14] on extremal indices. In higher dimensions results are scarce. There are, however, many multivariate models considered in literature with both serial dependence and marginal multivariate EVDs. Among the more important examples are multivariate linear models, see [5], random recursions, see [19] and stable processes, see [27].

In what follows we will make the assumption that our financial process has multivariate marginal distributions that are in the domain of attraction of an EVD with Fréchet one-dimensional marginals. We will also assume that the standard estimators of the tail index and of the spectral measure that are consistent for its associated process are also consistent for our financial process. Such consistency has been established for certain multivariate models; see Starica [28] for a discussion of GARCH type process.

2.2 Tails Under Aggregation

Financial risk managers are typically interested in the risk of large negative returns over horizons of length one day to one month. If the risk measure is quantile based, then it becomes crucial to estimate the tails of the return distribution accurately. Here there are two possibilities: The first is to include a longer history, spanning several years of data. The second one is to estimate the risk at a smaller return time interval and then scale this estimate up to match with the desired return time interval.

As markets change over time, structural breaks occur, new instruments

are introduced and others stop trading, implementing the first solution can be problematic. The second approach is also not problem free, especially when one enters the intra-day regime. A good example is the problem of exchange traded instruments on different exchanges with non-overlapping opening hours. E.g. defining the correlation between two such instruments for intra-daily returns is a non-trivial exercise. In the case of FX markets, however, this is not a problem as it is an over-the-counter market which is active around the clock.

The question that one now needs to answer is: What is the relationship between the risk estimated on short return time intervals with that based on longer return time intervals? As modern risk measures are based on the tail of the distribution the question can be reduced to: How do the tails behave under finite aggregation? Some empirical results exist for the one dimensional case [7], but here we focus on the multivariate case.

We start of by considering two random variables X_1 and X_2 and the tail of their sum $Y = X_1 + X_2$.

Theorem 2.1 *Let X_1 and X_2 be two regularly varying random variables in $\mathbf{R}^d$ with index α and spectral measures P_Θ^1 and P_Θ^2 and define $Y = X_1 + X_2$. Assume that*

$$\lim_{r \rightarrow \infty} P[\|X_1\| > r \|X_2\| > r] = 0. \quad (2)$$

Then Y is regularly varying and its spectral measure is a convex linear combination of P_Θ^1 and P_Θ^2 .

Proof: Observe first that by the nonnegativity of the random vectors

$$P[\|Y\| > r] \geq P[\|X_1\| > r] + P[\|X_2\| > r] - P[\|X_1\| > r, \|X_2\| > r],$$

and so by (2) and regular variation

$$\liminf_{r \rightarrow \infty} \frac{P[\|Y\| > r]}{P[\|X_1\| > r] + P[\|X_2\| > r]} \geq 1. \quad (3)$$

Next let F be any closed subset of the unit sphere. For any $0 < \epsilon < 1$ we have

$$\begin{aligned} P[\|Y\| > t, Y/\|Y\| \in F] &\leq P[\|X_1\| > (1 - \epsilon)t, \|X_2\| \leq \epsilon t, Y/\|Y\| \in F] \\ &+ P[\|X_2\| > (1 - \epsilon)t, \|X_1\| \leq \epsilon t, Y/\|Y\| \in F] \\ &+ P[\|X_1\| > \epsilon t, \|X_2\| > \epsilon t]. \end{aligned} \quad (4)$$

On the event $\{\|X_1\| > (1 - \epsilon)t, \|X_2\| \leq \epsilon t, Y/\|Y\| \in F\}$ we have

$$\left\| \frac{Y}{\|Y\|} - \frac{X_1}{\|X_1\|} \right\| \leq \frac{2\epsilon}{1 - \epsilon}$$

and, hence, $X_1/\|X_1\| \in F_{2\epsilon/(1-\epsilon)}$, where for a $\theta > 0$, F_θ is the closed θ -neighborhood of F . Therefore, by the regular variation of X_1 , the Portmanteau theorem (see Billingsley [4]) and (3), for any $x > 0$

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \frac{P[\|X_1\| > (1-\epsilon)tx, \|X_2\| \leq \epsilon tx, Y/\|Y\| \in F]}{P[\|Y\| > t]} \\ \leq & \limsup_{t \rightarrow \infty} \frac{P[\|X_1\| > (1-\epsilon)tx, X_1/\|X_1\| \in F_{2\epsilon/(1-\epsilon)}]}{P[\|X_1\| > t] + P[\|X_2\| > t]} \\ & \limsup_{t \rightarrow \infty} \frac{P[\|X_1\| > t] + P[\|X_2\| > t]}{P[\|Y\| > t]} \\ \leq & \limsup_{t \rightarrow \infty} \frac{P[\|X_1\| > (1-\epsilon)tx, X_1/\|X_1\| \in F_{2\epsilon/(1-\epsilon)}]}{P[\|X_1\| > t]} \\ & \limsup_{t \rightarrow \infty} \frac{P[\|X_1\| > t]}{P[\|X_1\| > t] + P[\|X_2\| > t]} \\ \leq & a_1(1 - \epsilon)^{-\alpha} x^{-\alpha} P_\Theta^1(F_{2\epsilon/(1-\epsilon)}), \end{aligned}$$

where $a_i = \lim_{r \rightarrow \infty} \frac{P[\|X_i\| > t]}{P[\|X_1\| > t] + P[\|X_2\| > t]}$, $i = 1, 2$. Since the second term in (4) can be treated in the same way, we let $\epsilon \rightarrow 0$ and use (2) and (3) to conclude that for any closed subset F of the unit sphere and $x > 0$

$$\limsup_{t \rightarrow \infty} \frac{P[\|Y\| > ts, Y/\|Y\| \in F]}{x^{-\alpha} P[\|Y\| > t]} \leq a_1 P_\Theta^1(F) + a_2 P_\Theta^2(F).$$

By the Portmanteau theorem this establishes regular variation of Y and the fact that its spectral measure is given by $a_1 P_\Theta^1 + a_2 P_\Theta^2$. Furthermore, it is clear that $0 \leq a_i \leq 1$ and $a_1 + a_2 = 1$. $\square$

The assumption (2) has to do with the dependence in the tail of the pair (X_1, X_2) . It says that the exceedence probability is zero. Exceedence probability is defined for two univariate positive random variables Z_1 and Z_2 as the limit

$$\lim_{r \rightarrow \infty} P[Z_1 > r | Z_2 > r]$$

whenever it exists.

This theorem can be extended to a sum of more than two variables in the following way.

Corollary 2.2 *Let $X_i, i = 1, 2, \dots, m$ be random variables such that pairwise they satisfy the conditions of Theorem 2.1. Define $Y = X_1 + \dots + X_m$, then Y is regularly varying with index α and its spectral measure is a convex linear combination of the spectral measures of X_i .*

Proof: By the above, the only thing that needs to be shown is that assumption (2) also holds for X_k and $X_i + X_j$ where i, j, k are different elements of $\{1, 2, \dots, m\}$. If this holds the proof follows by induction.

To this end observe that

$$\begin{aligned} P[\|X_k\| > r, \|X_i + X_j\| > r] &\leq P[\|X_k\| > r/2, \|X_i\| > r/2] \\ &+ P[\|X_k\| > r/2, \|X_j\| > r/2] \end{aligned}$$

and so the desired conclusion follows immediately from the regular variation and Theorem 2.1. $\square$

A simple consequence is that if the X_i 's all have the same spectral measure, then Y will have that same spectral measure. In particular, Y will have the same EVD type as each of the X_i 's.

Consider now the case when we have a stochastic process (X_i) in $\mathbf{R}^d$ such that the X_i 's all have the same distribution which is regularly varying with index α . Let (Y_i) denote the m -aggregated process, i.e. $Y_i = \sum_{k=im}^{(i+1)m-1} X_k$.

Theorem 2.3 *Let (X_i) and (Y_i) be as above and assume that*

$$\lim_{r \rightarrow \infty} P[\|X_i\| > r | \|X_j\| > r] = 0$$

for $i \neq j$. Then Y_i is regularly varying with index α and has the same spectral measure as X_i .

Proof: This is just an immediate consequence of Corollary 2.2 and the remark following it. $\square$

For the risk manager this means that, if the exceedance probabilities are zero, the EVD type, i.e. the tail index and the spectral measure, can be estimated from the high frequency time series (X_i) which has much more data and therefore allows a better estimate for the spectral measure and the tail index. The scale and location of the tail need to be estimate from the low frequency data, or alternatively, scaled up from those of the high frequency time series as is currently standard practice (scaling VaR with

$\sqrt{10}$ to obtain the ten day VaR from the one day VaR estimate). We will find the appropriate scaling empirically in Section 3.

In practice, there are limits to how small the return time intervals can be, i.e. how large m can be. For small return time intervals one might expect stronger serial dependence and assumptions such as (2) becomes less reliable, see e.g. [17].

3 Exploring the Empirical Tails

In order to verify the applicability of the theory outlined above to financial processes, it is important to assess how adequate its assumptions are. Among other things, we address the question of how well financial return distributions can be modelled by either elliptic or regularly varying distributions. Furthermore, we explore the serial dependence. We consider several return time intervals in order to gain an insight into how financial returns behave under aggregation.

Extremely large movements of financial markets are by definition rare and for a given dataset one can only expect a small portion of the data to represent extreme moves. For studying the extreme moves it is therefore imperative to have a large dataset to work with.

We consider 10 minutes to biweekly returns of the foreign exchange market. The returns are defined as

$$r_t = \frac{\log(x_t) - \log(x_{t-\Delta t})}{\Delta t}$$

where x_t is the geometric mean of the bid and ask price at time t and Δt is the return time interval. The foreign exchange market is active 24 hours a day, 7 days a week. We consider the 12 year period between January 1, 1987 and December 31, 1998. For 10 minute returns this gives us in excess of 630000 datapoints or in excess of 210000 for 30 minute returns. The time series from the market is of course not evenly spaced in time. Following [7], we used linear interpolation to obtain a regular time series. We study the four major FX rates USD/DEM, USD/CHF, USD/JPY and GBP/USD.

3.1 Spatial Dependence

We consider various features of the spatial dependence structure, i.e. dependence between different currencies. In the most general form this dependence structure could be captured by the copula [13], [18]. We are, however,

mostly interested in the dependence structure in the tail so our approach will be different. We first explore if the tails of the return distribution can be modelled by elliptic or regularly varying distributions. Then we estimate the exceedence probabilities and finally we estimate the spectral measure, which along with the tail index, completely determines the EVD type.

Normal distributions, or more generally elliptic distributions, have been a very popular choice for modelling financial returns. Elliptical distributions have some nice features, including closure under taking linear combinations and marginal distributions. While the tails of the normal distribution are too thin for most financial assets, it is still possible that some other elliptic distribution could fit financial returns better. Here we are mostly concerned with how well elliptic distributions capture the dependence structure in the tails.

Recall that a random variable X is said to be elliptic if there exists a constant vector μ and a positive definite matrix Σ such that the random variable $Y = \Sigma^{-1/2}(X - \mu)$ is spherically distributed, i.e its probability measure is invariant under rotations. The matrix Σ is then a constant multiple of the covariance matrix and μ is the mean, assuming that both exist. Note that if we define the sets

$$\Omega_s(\mu, \Sigma) = \{x | (x - \mu)\Sigma^{-1}(x - \mu) \geq s\}, \quad s \geq 0$$

then the conditioned random variable $X|_{\Omega_s}$ is also elliptic with the same μ and Σ . In particular X and $X|_{\Omega_s}$ both have the same correlation matrix. Therefore, if we estimate the correlation matrix of $X|_{\Omega_s}$ as a function of s it should be constant. In practice we fit an elliptic distribution to the entire data set, construct the appropriate ellipses, and then compute the correlations for the data lying outside of each ellipse. We now do this analysis for our data and the results are shown in Figure 1.

This Figure clearly indicates that the correlations increase as we get further into the tail. In particular, it is clear that elliptic distributions do not capture the correct dependence structure in the tails. We can now interpret the statement “financial assets are more strongly dependent when the market is in an excited state” by saying that the dependence in the tails of the return distribution is stronger than that indicated by elliptic distributions. The increase in tail conditional correlation is robust for all return time intervals. The increase is, however, more pronounced for shorter time intervals.

We have now established that the spatial dependence in the tails is not

Figure 1: This figure shows the correlation of the data lying outside an ellipse. The quantile indicates the fraction of data points lying inside the ellipse, the complement of Ω_s . The data used is 10 minute (solid), 30 minute (dotted), 2 hour (short-dashed) and daily (long-dashed) returns. All currencies are quoted against the US Dollar.

well captured by elliptical distributions fitted to the full distribution. Let us now consider regularly varying distributions. The main difference between these two classes is that in the former the dependence structure is not very flexible, but the radial behavior is. Regularly varying functions, on the other hand, have a more rigid radial behavior, but a flexible dependence structure.

It is generally accepted in finance and econometric literature that the univariate tails for a wide variety of financial assets have a power law decay and are therefore in the domain of attraction of the Fréchet EVD. Assuming regular variation, however, requires something more. Intuitively it requires that far out in the tail the probability density, assuming it exists, can be approximated by a product of the form $r^{-\alpha'} P_{\Theta}(\theta)$ where α' is independent of the angular coordinate θ . It is important, however, to realize that there may be parts of the unit sphere on which the spectral measure P_{Θ} vanishes. Such parts of the unit sphere, if exist, would corresponds to the directions in which the data has thinner tails, and this behavior is entirely consistent with regular variation.

For a bivariate process X let (R, Θ) be its representation in polar coordinates. The question is whether R , conditioned on $\Theta \in [\theta_0, \theta_1]$, has a constant tail index α for “many” $\theta_0 < \theta_1$. Recall that we expect to see a constant tail index α for those angles $[\theta_0, \theta_1]$ which carry a positive mass of the spectral measure. We check this by estimating the tail exponent of the conditional R process where we have split the unit circle into 12 equal sectors.

Figure 2 shows the tail exponent as a function of the average angle of the sector for three different horizons. The tail exponent is estimated by the Hill estimator on 1% of the data in each sector. In the case of daily returns we used overlapping intervals, with overlapping factor of 48. For 10 minute and 30 minute returns we used non-overlapping intervals. Using overlapping intervals has been shown to improve the statistics of an estimate [24], this applies in particular when conditioning as we do here. The conditioning on the tail of a sector reduces the amount of data in the estimate by a factor of 1200 which would leave us with less than 5 points for daily returns to estimate the tail from if not using overlapping intervals.

Figure 2 indicates that the tail index is relatively insensitive to the angle in the first and third quadrants, and it appears to be higher in the second and fourth quadrants (humps around $3\pi/4$ and $7\pi/4$). This is an indication that the spectral measures can carry little or no mass in the second and fourth quadrants.

Figure 2: Tail exponent as a function of angle. The plot shows results for 10 min. returns (solid), 30 min. returns (dotted) and daily returns (dashed).

Another method for measuring how probability mass is distributed far out in the tails is to estimate the exceedence probabilities. For two univariate random variables X and Y let x_q and y_q denote the q -th quantile of X and Y respectively. By the positive symmetric exceedence probabilities we mean the following limit

$$\lim_{q \rightarrow 1} P[Y \geq y_q | X \geq x_q]$$

given that it exists. The asymmetric exceedence probabilities are defined in a similar way for $(X, -Y)$ and $(-X, Y)$ and the negative symmetric exceedence probabilities are defined for $(-X, -Y)$. If these limits are all zero we say that X and Y are asymptotically independent. It is known that various standard distributions, including the nondegenerate normal and Student-t distributions, are asymptotically independent.

Figure 3: Symmetric exceedence probabilities as a function of the quantile. The data used is 10 minute (solid), 30 minute (dotted), 2 hour (short-dashed) and daily (long-dashed) returns. All currencies are quoted against the US Dollar.

Figure 3 presents the symmetric exceedence probabilities as a function of the quantile. From these graphs it is clear that the limit is greater than zero and hence there is dependence in the tails of these processes. The increased exceedence probability for longer return time intervals is likely related to the Epps effect [16] which states that correlations decrease as the return time interval becomes smaller in financial markets. Also notice that the exceedence probability for daily returns and bihourly returns are quite close, whereas those for 10 minute returns are significantly different.

Figure 4 presents the asymmetric exceedence probabilities. Here the limit is clearly zero. It is also striking how similar the exceedences are for 30 minute to daily returns.

Figure 4: Asymmetric exceedence probabilities as a function of the quantile. The data used is 10 minute (solid), 30 minute (dotted), 2 hour (short-dashed) and daily (long-dashed) returns. All currencies are quoted against the US Dollar.

The exceedence probabilities seem to be very robust for various return

time intervals. One would therefore expect to have positive symmetric exceedence probabilities and zero asymmetric exceedence probabilities for returns over longer time intervals than one day.

Recall that the spectral measure captures completely the dependence structure of the EVD. We estimate it by estimating the density of Θ conditional on R being in the 99-th quantile. For simplicity, we again focus on the bivariate setting.

Figure 5: The spectral measure as a function of angle. The data used is 10 minute (solid), 30 minute (dotted), 2 hour (short-dashed) and daily (long-dashed) returns. All currencies are quoted against the US Dollar.

Figure 5 presents the spectral measure as a function of angle for the different assets. If the tails of the asset returns show positive dependence then we would expect probability mass to be more concentrated in the first and the third quadrant. This is indeed the case and moreover the plots clearly indicate that the probability mass in the second and the fourth quadrants is

very small. This phenomenon is almost extreme with European currencies. This is consistent with our results for the exceedence probabilities.

It is very striking how similar the spectral measures are and these plots indicate that the spectral measure for long return intervals can be estimated from short return time interval data to a high degree of accuracy. These empirical results are consistent with the result of Theorem 2.3.

3.2 Serial Dependence

So far we have only considered the marginal distributions of the process in our analysis. Below we explore the dependence of the time series (X_t, X_{t-1}) . This yields important information about clustering of the extremes and supports our assumption (2) of Theorem 2.1.

Figure 6 shows estimates of the spectral density of a lagged time series versus a non-lagged time series for our four currency pairs. When the spectral measure is concentrated on the horizontal and vertical axis (shown as dotted vertical lines) it indicates that the two variables are independent in the tails. The figure clearly indicates that as we move from 10 minute returns to bihourly returns or even 6 hour returns the mass gets more focused around the axis and the dependence decreases. For longer return time intervals, in particular daily returns, we have less data and the estimate of the spectral measure becomes less reliable.

We also estimate the symmetric and asymmetric exceedence probabilities for the lagged versus non-lagged time series. The symmetric ones are shown in Figure 7 and the asymmetric ones in Figure 8.

These figures indicate that there is a strictly positive, but small, exceedence probability for the shorter return time intervals, 10 minute and 30 minute. For bihourly returns, we can argue that the exceedence probability is zero for both the symmetric and the asymmetric case. One might even argue that this is also the case for hourly returns in the asymmetric case. For the longer returns we again have less data and the results are less reliable. One would expect that the exceedence probability would decrease with time and hence that exceedence probabilities would be zero for return time intervals longer than two hours. Zero exceedence probabilities indicate that assumption (2) of Theorem 2.1 is valid.

Our next step is to consider the extent to which our estimates of the tail index and of the spectral measure are affected by the serial dependence in the data. Here we choose to focus on 30 minute returns as for longer return time intervals we have less data.

Figure 6: The spectral measure for time series X versus time series X lagged by one return time interval. The solid line represents $X = \text{USD/DEM}$, the dotted line represents $X = \text{USD/CHF}$, the short-dashed line represents $X = \text{USD/JPY}$ and the long-dashed line represents $X = \text{GBP/USD}$.

Figure 7: The positive symmetric exceedance probability for time series X versus time series X lagged by one return time interval. The solid line represents $X = \text{USD/DEM}$, the dotted line represents $X = \text{USD/CHF}$, the short-dashed line represents $X = \text{USD/JPY}$ and the long-dashed line represents $X = \text{GBP/USD}$.

Figure 8: The asymmetric exceedance probability for time series X versus time series X lagged by one return time interval. The solid line represents $X = \text{USD/DEM}$, the dotted line represents $X = \text{USD/CHF}$, the short-dashed line represents $X = \text{USD/JPY}$ and the long-dashed line represents $X = \text{GBP/USD}$.

To minimize the effect of the serial dependence on the estimates, we follow Zachary *et al.* [30] and split our sample into subsamples of length 100. We then choose the single largest observation from each subsample. We regard the resulting set of observations as a new sample. The spectral measure is then estimated from all of this new sample. Similarly, the tail index is estimated from this new sample, but keep in mind that the Hill estimator we are employ uses only a subsample of it, containing the m largest observations. If the subsamples are long enough, then the probability of two observations used for the estimate being close together is small and we have obtained a reduced declustered sample.

Figure 9: The tail index of 30 minute returns as estimated by the Hill estimator as a function of order statistic. The solid line represents the full sample estimate for (X_i) and the dotted line represents the reduced sample estimate for $(\tilde{X}_i)$.

Figure 9 presents the Hill estimate of the tail index as a function of the order statistic. The tail index as estimated for the declustered sample

appears to be systematically smaller than that estimated for (X_i) for larger order statistics. Beside the effect of clustering, this may be due to the fact that largest observations were selected in each subsample of length 100. As the order statistic gets smaller the two estimates converge. This could be either because the very largest extremes occur individually and not in clusters or because the subsamples are not long enough to decluster the very largest extremes. In any case, bearing in mind the difficulties of estimating tail indices [7], the difference between the two estimated curves is rather small.

Figure 10: The spectral measure of 30 minute returns as a function of angle. The solid line represents the full sample estimate for (X_i) and the dotted line represents the reduced sample estimate for $(\tilde{X}_i)$.

Figure 10 presents the spectral measure estimate for both (X_i) and the declustered sample. The match here is strikingly good and once more the spectral measure is shown to be a very robust feature of the marginal return

distributions.

From these results on the tail index and the spectral measure we can conclude that not taking clustering into account is reasonable, even for 30 minute return time intervals.

4 Risk Management

Our empirical results of Section 3 show that the assumptions of the theory in Section 2 are satisfied. Furthermore, the robustness of the spectral measure under aggregation supports the results of the theory. We now turn our attention to applying this theory to risk management. We first address the issue of estimating risk for horizons which are longer than those we have analyzed until now. Secondly, we consider risk minimization for those same horizons. Our goal here is to show that if a risk measure only depends on the EVD of the returns, then the risk can be estimated and minimized at a shorter return time interval where more of reliable data are available.

Estimating risk using methods from univariate extreme value theory has been a popular topic in the econometrics literature in recent years, see e.g. [7], [8] and [15]. This interest in extreme value theory has gone hand in hand with the popularity of quantile based risk estimation methods. Most authors have here focused on Value-at-Risk (VaR) as the appropriate measure of risk. The VaR can be defined for a given threshold level γ as the γ -th quantile of the portfolio return distribution. Recent examples [3], however, have been provided which show that VaR may not be as good a measure of risk as was originally hoped for. In particular, the VaR is not always subadditive which means that diversification does not always decrease the risk as measured by VaR. This can produce catastrophic results if the risk estimates are then used for portfolio optimization. An alternative risk measure which overcomes these problems has been proposed, namely the expected shortfall (ES) also known as the conditional VaR or the mean excess. ES can be defined for a given threshold γ as the average loss given that the loss exceeds the γ -th quantile.

Some efforts have been made to characterize what constitutes a good measure of risk. A reasonable set of criteria has been proposed by Artzner *et al.* [2], [3], who defined the notion of a *coherent measure* of risk. If we let $\lambda(X)$ be the measure of the risk associated with holding the instrument X and let r denote the risk free interest rate, then the four conditions can be expressed as:

Translation invariance: $\lambda(X + \eta \cdot r) = \lambda(X)$

Subadditivity: $\lambda(X_1 + X_2) \leq \lambda(X_1) + \lambda(X_2)$

Positive homogeneity: for all $s \geq 0$, $\lambda(sX) = s\lambda(X)$

Monotonicity: for $X_1 \leq X_2$, $\lambda(X_1) \leq \lambda(X_2)$

So far we have shown, both theoretically and empirically, that the EVD type of hourly or bihourly returns is the same as that of daily returns. Given the robustness of the spectral measure, we can assume that this can be extrapolated so that weekly, biweekly or even monthly returns have the same EVD type. Equivalently, we can assume that these returns are all regularly varying with the same tail index and the same spectral measure.

What still is missing is how the tails scale under aggregation. Current practice is to assume that VaR scales with the square root of the return time interval. Some evidence have been provided against this practice [7], [9]. Figure 11 show VaR as a function of return time interval on a log-log scale and Figure 12 shows the same for ES.

We have fitted straight lines through the points in these figures and the fits appear to be quite good. This indicates that the risk as estimated by VaR or ES does indeed scale with a power law in return time interval. In other words, if we let $\text{VaR}(\Delta t)$ denote the Value-at-Risk as estimated for horizon Δt , measured in days, then we have the following scaling

$$\text{VaR}(\Delta t) = \text{VaR}(1)(\Delta t)^\kappa$$

where κ is called the scaling exponent. The scaling is defined in an identical manner for the ES. The scaling exponent we observe is not 0.5 as for Brownian motion, but slightly smaller. The results are shown in Table 1. These results are related to the scaling results of [23] for $E[|r|^p]$ for various p .

	DEM	CHF	JPY	GBP	mean
VaR	0.474	0.456	0.465	0.487	0.47
ES	0.449	0.441	0.441	0.461	0.45

Table 1: The scaling exponent κ for VaR and ES. All currencies are quoted against the USD.

Consider now the problem of allocating capital between two foreign currencies so that the risk is minimized from the standpoint of a US investor.

Figure 11: VaR as a function of return time interval on a log-log scale. Average scaling exponent is $\kappa = 0.47$

Figure 12: ES as a function of return time interval on a log-log scale. Average scaling exponent is $\kappa = 0.45$

This involves finding the best allocation parameter $\lambda \in [0,1]$ such that the fraction of the capital allocated to the first currency is λ and the fraction allocated to the second one is $1 - \lambda$. As a result of our analysis, if the risk measure only depends on the EVD of the return distribution, then this minimization can be done based on data from short return time intervals even though the horizon in question is much longer.

Figure 13: VaR at the 99-th quantile for a 2 week horizon as estimated by hourly data (solid), daily data (dotted) and biweekly data (dashed)

Figure 13 shows the VaR for a two week horizon as a function of the allocation parameter. The VaR is estimated by hourly data, daily data and biweekly data. First notice that all the curves are at the same level, indicating that the level of risk as given by the VaR is accurately estimated by the hourly data. Secondly, the curves with hourly and daily data have similar shapes.

Figure 14 shows the same thing for the ES. Here the hourly and daily

Figure 14: ES at the 99-th quantile for a 2 week horizon as estimated by hourly data (solid), daily data (dotted) and biweekly data (dashed)

data curves are even closer than before and both are fairly smooth. Again the general level of risk is correctly estimated by the hourly data.

The biweekly curves in both figures are based on too few datapoints for the estimate of its shape to be reliable. Here we are using 12 years of historical data. A more realistic exercise would be to use 1 year of data and then the quality of the daily data curve would have deteriorated to the same level as that of the biweekly curve currently. This clearly shows that in order to do any risk minimization at the 99-th quantile level of risk one needs to use data of higher frequency than daily.

As a last point we note that the curves for hourly and daily data for ES are smoother than those for VaR, hence these curves lend themselves better to risk minimization by an optimization algorithm. In any case, doing any kind of risk minimization using VaR as a measure of risk is dangerous as VaR is not capable of detecting concentration of risk, i.e. it does not always decrease with diversification.

5 Conclusion

The common sense knowledge of financial practitioners that financial asset returns are “more dependent when the market is excited” is not well explained by elliptical distributions. We show that regularly varying distributions are better suited for capturing the dependence structure of the tails of foreign exchange return distributions than elliptical distributions.

In this paper we explore both the spatial (between assets) and serial dependence structure of extreme movements in the foreign exchange market. Our study includes a range of return time lengths to provide a comprehensive view of how this dependence behaves under aggregation. In particular, we estimate the spatial and serial spectral measures and show that the former are remarkably robust under aggregation as well as being insensitive to serial dependence of the time series. We estimate symmetric and asymmetric exceedence probabilities for spatial dependence which are also robust with respect to aggregation. Furthermore, we show that for serial dependence the exceedence probabilities are positive, but small, for shorter return time intervals but zero for return time intervals longer than two hours.

In tandem with our empirical studies we have developed theory which predicts that the extreme value distribution of hourly or bihourly returns is the same as that of daily, weekly or even monthly returns. The assumptions of this theory have been empirically verified. Together with our empiri-

cal results this, clearly indicates that high frequency data can considerably increase quality of estimates of extreme events.

Lastly, we have applied our results to risk management. Here we have developed methods for estimating risk over relatively long horizons based on hourly or bihourly data. These methods allow for better statistical estimates than commonly used, which is crucial when engaging in any form of risk minimization.

References

- [1] R.J. Adler, R.E. Feldman, and M.S. Taqqu, editors. *A Practical Guide to Heavy Tails*. Birkhäuser, Boston, 1998.
- [2] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Thinking coherently. *Risk*, 10(11), November 1997.
- [3] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. Preprint, April 8 1998.
- [4] P. Billingsley. *Convergence of Probability Measures*. Second edition. Wiley, New York, 1999.
- [5] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 1987.
- [6] B.N. Cheng and S.T. Rachev. Multivariate stable securities in financial markets. Technical report, University of California, Santa Barbara, 1999.
- [7] Michel M. Dacorogna, Olivier V. Pictet, Ulrich A. Müller, and Casper G. de Vries. The distribution of extremal foreign exchange rate returns in extremely large data sets. Internal document UAM.1992-10-22, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland, May 20 1994.
- [8] J. Danielsson and C.G. de Vries. Value-at-Risk and extreme returns. Technical report, London School of Economics, 1997.
- [9] J. Danielsson, P. Hartmann, and C.G. de Vries. The cost of conservatism: Extreme returns, Value-at-Risk and the Basle 'multiplication factor'. *Risk*, 1 1998.

- [10] R.A. Davis, T. Mikosch, and B. Basrak. Sample ACF of multivariate stochastic recurrence equations with applications to GARCH. <http://www.cs.rug.nl/~eke/iwi/preprints/>, 1999.
- [11] L. de Haan, S.I. Resnick, H. Rootzén, and C.G. de Vries. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications*, 32:213–224, 1989.
- [12] Laurens de Haan and John de Ronde. Sea and wind: Multivariate extremes at work. *Extremes*, 1(1):7–45, 1 1998.
- [13] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependency in risk management: Properties and pitfalls. Technical report, ETH Zurich, 1998.
- [14] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events*, volume 33 of *Applications of Mathematics Stochastic Modelling and Applied Probability*. Springer, Berlin, 1997.
- [15] S. Emmer and C. Klüppelberg. VaR - ein Mass für das extreme Risiko. *Solutions*, 1:53–63, 1998.
- [16] T.W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298, 1 1979.
- [17] Dominique M. Guillaume, Michel M. Dacorogna, Rakhal D. Davé, Ulrich A. Müller, Richard B. Olsen, and Olivier V. Pictet. From the bird’s eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1:95–129, 3 1997.
- [18] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman and Hall, London, 1997.
- [19] Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1 1973.
- [20] C. Klüppelberg and A. May. The dependence function for bivariate extreme value distributions - a systematic approach. *Technical University of Munich*, 1998.
- [21] M.R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York Berlin, 4 1983.

- [22] F. Longin and B. Slonik. Correlation structure of international equity markets during extremely volatile periods. Technical Report CR 646/1998, Chambre de Commerce et D'Industrie de Paris, 1998.
- [23] U. A. Müller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance*, 14:1189–1208, 12 1990.
- [24] Ulrich A. Müller. Statistics of variables observed over overlapping intervals. Internal document UAM.1993-06-18, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland, June 18 1993.
- [25] Ulrich A. Müller, Michel M. Dacorogna, and Olivier V. Pictet. Heavy tails in high-frequency financial data. *published in the book "A practical guide to heavy tails: Statistical Techniques for Analysing Heavy Tailed Distributions", edited by Robert J. Adler, Raisa E. Feldman and Murad S. Taqqu and published by Birkhäuser, Boston 1998, pages 55–77, 8 1998.*
- [26] S.I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York, 1987.
- [27] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York, 1994.
- [28] Catalin Starica. Estimation of the extreme value distribution for constant conditional correlation models. *Proceedings of the HFDF-II Conference in Zurich*, 2:1–30, 3 1998.
- [29] Stefan Straetmans. *Extreme financial returns and their comovements*. PhD thesis, Erasmuss University Rotterdam, May 1998.
- [30] S. Zachary, G. Feld, G. Ward, and J. Wolfram. Multivariate extrapolation in the offshore environment. *Applied Ocean Research*, 20:273–295, 1998.