

Financial Risk and Heavy Tails

Brendan O. Bradley and Murad S. Taqqu
Boston University

August 28, 2001

Abstract

It is of great importance for those in charge of managing risk to understand how financial asset returns are distributed. Practitioners often assume for convenience that the distribution is normal. Since the 1960s, however, empirical evidence has led many to reject this assumption in favor of various heavy-tailed alternatives. In a heavy-tailed distribution the likelihood that one encounters significant deviations from the mean is much greater than in the case of the normal distribution. It is now commonly accepted that financial asset returns are, in fact, heavy-tailed. The goal of this survey is to examine how these heavy tails affect several aspects of financial portfolio theory and risk management. We describe some of the methods that one can use to deal with heavy tails and we illustrate them using the *NASDAQ* composite index.

Contents

1	Introduction	4
2	Historical Perspective	5
2.1	Risk and Utility	5
2.2	Markowitz Mean-Variance Portfolio Theory	5
2.3	CAPM and APT	6
2.4	Empirical Evidence	9
3	Value at Risk	10
3.1	Computation of VaR	13
3.1.1	Historical Simulation VaR	13
3.1.2	Parametric VaR	14
3.1.3	Monte Carlo VaR	16
3.2	Parameter Estimation	17
3.2.1	Historical Volatility	17
3.2.2	ARCH/GARCH Volatilities	19
3.2.3	Implied Volatilities	22
3.2.4	Extreme Value Theory	22
4	Risk Measures	23
4.1	Coherent Risk Measures	23
4.2	Expected Shortfall	24
5	Portfolios and Dependence	25
5.1	Copulas	26
5.2	Measures of Dependence	30
5.2.1	Linear Correlation	31
5.2.2	Rank Correlation	31
5.2.3	Comonotonicity	32
5.2.4	Tail Dependence	33
5.3	Elliptical Distributions	35
6	Univariate Extreme Value Theory	39
6.1	Limit Law for Maxima	40
6.2	Block Maxima Method	41
6.3	Using the Block Maxima Method for Stress Testing	43
6.4	Peaks Over Threshold Method	44
6.4.1	Semiparametric Approach	44
6.4.2	Fully Parametric Approach	46
6.4.3	Numerical Illustration	50
6.4.4	A GARCH-EVT model for Risk	51

7	Stable Paretian Models	54
7.1	Stable Portfolio Theory	55
7.2	Stable Asset Pricing	57

1 Introduction

Financial theory has long recognized the interaction of risk and reward. The seminal work of Markowitz [Mar52] made explicit the trade-off of risk and reward in the context of a portfolio of financial assets. Others such as Sharpe [Sha64], Lintner [Lin65], and Ross [Ros76], have used equilibrium arguments to develop asset pricing models such as the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), relating the expected return of an asset to other risk factors. A common theme of these models is the assumption of normally distributed returns. Even the classic Black and Scholes option pricing theory [BS73] assumes that the return distribution of the underlying asset is normal. The problem with these models is that they do not always comport with the empirical evidence. Financial asset returns often possess distributions with tails heavier than those of the normal distribution. As early as 1963, Mandelbrot [Man63] recognized the heavy-tailed, highly peaked nature of certain financial time series. Since that time many models have been proposed to model heavy-tailed returns of financial assets.

The implication that returns of financial assets have a heavy-tailed distribution may be profound to a risk manager in a financial institution. For example, 3σ events may occur with a much larger probability when the return distribution is heavy-tailed than when it is normal. Quantile based measures of risk, such as value at risk, may also be drastically different if calculated for a heavy-tailed distribution. This is especially true for the highest quantiles of the distribution associated with *very rare* but *very damaging* adverse market movements.

This paper serves as a review of the literature. In Section 2, we examine financial risk from an historical perspective. We review risk in the context of the mean-variance portfolio theory, CAPM and the APT, and briefly discuss the validity of their assumption of normality. Section 3 introduces the popular risk measure called *value at risk* (VaR). The computation of VaR often involves estimating a scale parameter of a distribution. This scale parameter is usually the volatility of the underlying asset. It is sometimes regarded as constant, but it can also be made to depend on the previous observations as in the popular class of ARCH/GARCH models.

In Section 4, we discuss the validity of several risk measures by reviewing a proposed set of properties suggested by Artzner, Delbean, Eber and Heath [ADEH99] that any sensible risk measure should satisfy. Measures satisfying these properties are said to be *coherent*. The popular measure VaR is, in general, not coherent, but the *expected shortfall* measure is. The expected shortfall, in addition to being coherent, gives information on the expected size of a large loss. Such information is of great interest to the risk manager.

In Section 5, we return to risk, portfolios and dependence. *Copulas* are introduced as a tool for specifying the dependence structure of a multivariate distribution separately from the univariate marginal distributions. Different measures of dependence are discussed including *rank correlations* and *tail dependence*. Since the use of linear correlation in finance is ubiquitous, we introduce the class of *elliptical distributions*. Linear correlation is shown to be the canonical measure of dependence for this class of multivariate distributions and the standard tools of risk management and portfolio theory apply.

Since the risk manager is concerned with extreme market movements we introduce *extreme value theory* (EVT) in Section 6. We review the fundamentals of EVT and argue that it shows great promise in quantifying risk associated with heavy-tailed distributions. Lastly, in Section 7, we examine the use of *stable distributions* in finance. We reformulate the mean-variance portfolio theory of Markowitz and the CAPM in the context of the multivariate stable distribution.

2 Historical Perspective

2.1 Risk and Utility

Perhaps the most cherished tenet of modern day financial theory is the trade-off between risk and return. This, however, was not always the case, as Bernstein's [Ber96] narrative on risk indicates. In fact, investment decisions used to be based primarily on expected return. The higher the expected return, the better the investment. Risk considerations were involved in the investment decision process, but only in a qualitative way, *stocks are more risky than bonds*, for example. Thus any investor considering only the expected payoff $\mathbb{E}X$ of a game (investment) would, in practice, be willing to pay a fee equal to $\mathbb{E}X$ for the right to play.

The practice of basing investment decisions solely on expected return is problematic, however. Consider the game known today as the *Saint Petersburg Paradox*, introduced in 1728 by Nicholas Bernoulli. The game involves flipping a fair coin and receiving a payoff of 2^{n-1} roubles¹ if the first head appears on the n^{th} toss of the coin. The longer tails appears, the larger the payoff. While in this game the expected payoff is infinite, no one would be willing to wager an infinite sum to play, hence the paradox. Investment decisions cannot be made on the basis of expected return alone.

Daniel Bernoulli, Nicholas' cousin, proposed a solution to the paradox ten years later. He believed that, instead of trying to maximize their expected wealth, investors want to maximize their expected *utility* of wealth. The notion of utility is now widespread in economics². A utility function $U : \mathbb{R} \rightarrow \mathbb{R}$ indicates how desirable is a quantity of wealth W . One generally agrees that the utility function U should have the following properties:

1. U is continuous and differentiable over some domain D .
2. $U'(W) > 0$ for all $W \in D$, meaning investors prefer more wealth to less.
3. $U''(W) < 0$ for all $W \in D$, meaning investors are risk averse. Each additional dollar of wealth adds less to the investors utility when wealth is large than when wealth is small.

In other words, U is smooth and concave over D . An investor can use his utility function to express his level of risk aversion.

2.2 Markowitz Mean-Variance Portfolio Theory

In 1952, while a graduate student at the University of Chicago, Harry Markowitz [Mar52] produced his seminal work on portfolio theory connecting risk and reward. He defined the reward of the portfolio as the expected return and the risk as its standard deviation or variance³. Since the expectation operator is linear, the portfolio's expected return is simply given by the weighted sum of the individual assets' expected returns. The variance operator, however, is not linear. This means that the risk of a portfolio, as measured by the variance, is not equal to the weighted sum of risks of the individual assets. This provides a way to quantify the benefits of diversification.

We briefly describe Markowitz' theory in its classical setting where we assume that the assets distribution is multivariate normal. We will relax this assumption in the sequel. For example, in

¹In fact, it was *ducats* [Ber96].

²For introductions to utility theory see for example Ingersoll [Ing87] or Huang and Litzenberger [HL88].

³In practice, one minimizes the variance, but it is convenient to view risk as measured by the standard deviation.

Section 5.3, we will suppose that the distribution is elliptical and, in Section 7.1, that it is an infinite variance stable distribution.

Consider a universe with n risky assets with random rates of return $\mathbf{X} = (X_1, \dots, X_n)$, with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, covariance matrix $\boldsymbol{\Sigma}$ and portfolio weights $\mathbf{w} = (w_1, \dots, w_n)$. If \mathbf{X} is assumed to have a multivariate normal distribution $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the return distribution of the portfolio $X_p = \mathbf{w}^T \mathbf{X}$ is also normally distributed, $X_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$ where $\mu_p = \mathbf{w}^T \boldsymbol{\mu}$ and $\sigma_p^2 = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$. The problem is to find the portfolio of minimum variance that achieves a minimum level a of expected return:

$$\begin{aligned} & \min_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}, \\ \text{such that } & \mathbf{w}^T \boldsymbol{\mu} \geq a, \\ & \mathbf{e}^T \mathbf{w} = 1. \end{aligned} \tag{1}$$

Here $\mathbf{e} = (1, \dots, 1)$ and T denotes a transpose. The last condition in (1), $\mathbf{e}^T \mathbf{w} = \sum_{i=1}^n w_i = 1$, indicates that the portfolio is fully invested. Additional restrictions are usually added on the weights⁴ and the problem is generally solved through quadratic programming. By varying the minimum level a of expected return, a set of portfolios X_p is chosen, each of which is optimal in the sense that an investor cannot achieve a greater expected return, $\mu_p = \mathbb{E}X_p$, without increasing his risk, σ_p . The set of optimal portfolios corresponds to a convex curve $(\sigma_p, \mathbb{E}X_p)$ called the *efficient frontier*. Any rational investor making decisions based only on the mean and variance of the distribution of returns of a portfolio would only choose to own portfolios on this efficient frontier. The specific portfolio he chooses depends on his level of risk aversion⁵. If the universe of assets also includes a *risk-free* asset which the investor may borrow and lend without constraint, then the optimal portfolio is a linear combination of the risk-free asset r and a certain risky portfolio X_R on the efficient frontier. As shown in Figure 1, this line is tangent to the convex risky asset efficient frontier at the point $(\sigma_R, \mathbb{E}X_R)$. The risky portfolio therefore maximizes the slope of this linear combination,

$$\max_{\mathbf{w}} \frac{\mathbb{E}(X_R) - r}{\sigma_{X_R}}. \tag{2}$$

Again, the specific weights given to the risk-free and risky assets depend on the individual investors level of risk aversion.

2.3 CAPM and APT

The mean-variance portfolio theory of Markowitz describes the construction of an optimal portfolio, in the mean-variance sense, for an individual investor. It requires only estimates for each asset mean

⁴For example, $w_i \geq 0$, in other words no *short* selling. Without the additional constraints, the problem can be solved by using linear programming.

⁵One can reconcile maximizing expected utility with the mean-variance portfolio theory of Markowitz, but one has to assume either a quadratic utility function or that returns are multivariate normal or, more generally, elliptical. (Elliptical distributions are introduced in Section 5.3). For example, if returns are multivariate normal and if X_{p_1} and X_{p_2} are the returns of two linear portfolios with the same expected return, then for all utility functions U with properties listed in Section 2.1,

$$\mathbb{E}U(X_{p_1}) \geq \mathbb{E}U(X_{p_2}) \quad \text{if and only if} \quad \sigma_{p_1}^2 \leq \sigma_{p_2}^2.$$

See for example [Ing87].

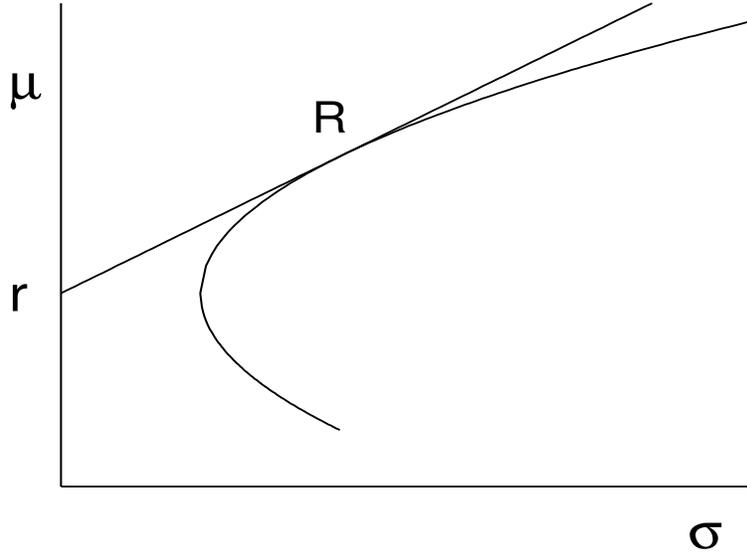


Figure 1: The efficient frontier (σ_p, μ_p) . In the case when only risky assets R are available, the frontier traces out a convex curve in risk-return space. The inclusion of a risk-free asset r , has a profound effect on the efficient set. In this case, all efficient portfolios will consist of linear combinations of r and some risky portfolio R , where (σ_R, μ_R) lies on the efficient frontier.

return, and the covariance between assets⁶. If all investors act in a way consistent with Markowitz' theory, then under additional assumptions, one will be able to learn something about the trade-off between risk and return in a market in equilibrium⁷. This is what the CAPM does.

The *capital asset pricing model* (CAPM) is an equilibrium pricing model (see Sharpe [Sha64] and Lintner [Lin65]) which relates the expected return of an asset to the risk-free return, to the *market's* expected return and to the covariance between the market and the asset. In addition to assuming that market participants use the mean-variance framework, the model makes two additional major assumptions. First, the market is assumed *frictionless*. This means that securities are infinitely divisible, there exist no transaction costs, no taxes, and there are no trading restrictions. Second, the investors beliefs are *homogeneous*. This means investors agree on mean returns and covariances for all assets in the market.

The efficient frontier in Figure 1 depended on the investors' belief. Under the CAPM assumptions, since all investors assume the same expected return and covariances for all assets in the market, they all have the same (risky) efficient frontier. However, the individual investors choice of the optimal risky portfolio still depends on the investors own level of risk aversion. Additionally, with the inclusion of a risk-free asset, we saw that the investors portfolios become dramatically more simple. Each investor can own only two assets: the risk-free asset and an optimal risky portfolio, with the relative weights depending on the investors appetite for risk. But since each investor holds the same optimal portfolio

⁶For a universe of n assets it is necessary to compute $n(n-1)/2 + n$ covariances. This means that if the universe under consideration consists of $n = 1000$ assets, it is necessary to estimate over 500 000 covariances.

⁷By market equilibrium, we mean a market place where security prices are set so that supply equals demand.

of risky assets, and since the market is assumed to be in equilibrium, this optimal risky portfolio must be the *market* portfolio. Thus Figure 1 applies with $R = M$, where M denotes the market portfolio. M consists of all risky assets held in proportion to their overall market capitalization. Letting X_M denote the return on the market portfolio, X_i denote the return of asset i , and r denote the risk-free return, the CAPM establishes the following relationship:

$$\mathbb{E}(X_i - r) = \beta_i \mathbb{E}(X_M - r) \quad (3)$$

where

$$\beta_i = \frac{\text{Cov}(X_i, X_M)}{\text{Var}X_M}. \quad (4)$$

The CAPM thus relates in a linear way the expected premium $\mathbb{E}X_i - r$ of holding the risky asset i over the risk-free asset to the expected premium $\mathbb{E}X_M - r$ of holding the market portfolio over the risk-free asset. The constant of proportionality is the asset's *beta*. The coefficient β_i is a measure of asset i 's sensitivity to the market portfolio. The expected premium for asset i is greater than that of the market if $\beta_i > 1$ and less if $\beta_i < 1$. But if $\beta_i > 1$, then the risk will be greater. Indeed, if we assume that

$$X_i - r = \beta_i(X_M - r) + \epsilon_i, \quad (5)$$

where ϵ_i is such that $\mathbb{E}\epsilon_i = 0$ and $\text{Cov}(\epsilon_i, X_M) = 0$, then we have (3) and

$$\sigma_{X_i}^2 = \beta_i^2 \sigma_{X_M}^2 + \sigma_{\epsilon_i}^2. \quad (6)$$

Equation (5) is often known as a *single factor model* for asset returns. Notice from (6) that the asset's risk is the sum of two terms, the *systematic* or *market* risk $\beta_i^2 \sigma_{X_M}^2$ and the *unsystematic* or *residual risk* $\sigma_{\epsilon_i}^2$. For a portfolio X_p with weights $\mathbf{w} = (w_1, \dots, w_n)$, one gets similarly $\sigma_{X_p}^2 = \beta_p^2 \sigma_{X_M}^2 + \sigma_{\epsilon_p}^2$ where $\beta_p = \sum_{i=1}^n w_i \beta_i$. If one additionally assumes that $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$ then the residual risk is

$$\sigma_{\epsilon_p}^2 = \sum_{i=1}^n w_i^2 \sigma_{\epsilon_i}^2. \quad (7)$$

It is bounded by c/n for some constant c , if for example, $w_i = 1/n$, and hence the portfolio's residual risk can be greatly reduced by diversification. The investor, for example, is only rewarded for bearing systematic or market risk, that is, he can expect a higher return than the market only by holding a portfolio which is riskier ($\beta_p > 1$) than the market.

In the CAPM, all assets are exposed to a single *common* source of randomness, namely the market. The *arbitrage pricing theory* (APT) model, due to Ross [Ros76], is a generalization of the CAPM in which assets are exposed to a larger number of common sources of randomness. The APT differs from the CAPM in that the mean-variance framework that led to (5) is now replaced by the assumption of a *multifactor* model

$$X_i = \alpha_i + \beta_{i1}f_1 + \dots + \beta_{ik}f_k + \epsilon_i \quad (8)$$

for generating security returns. All assets are exposed to the k sources of randomness f_j , $j = 1, \dots, k$, called factors. Additionally, each asset i is exposed to its own specific source of randomness ϵ_i . The equilibrium argument used in the CAPM led to the central result (3). In the APT, the equilibrium assumption takes a slightly different form, namely, one assumes that the market is free of arbitrage.

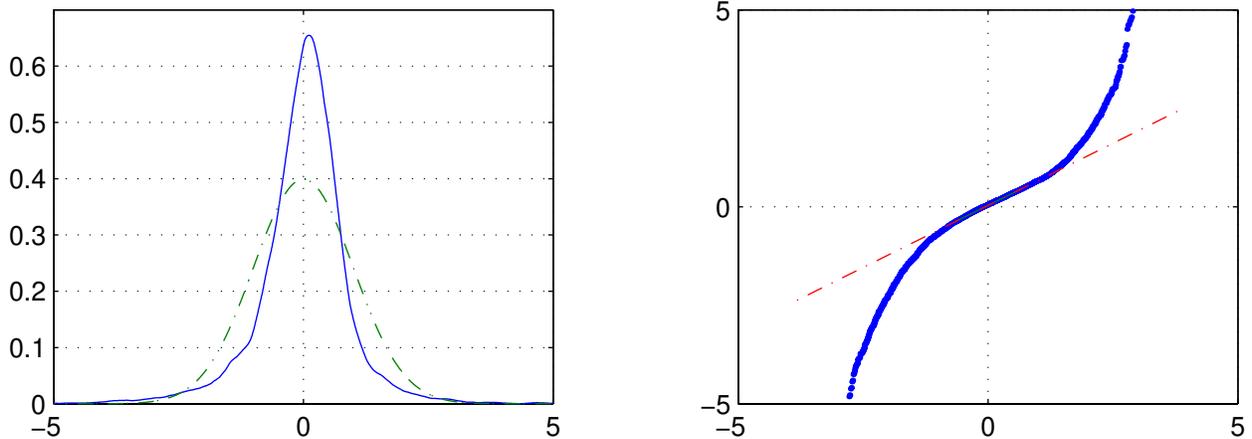


Figure 2: Left: Empirical probability density function (pdf) for NASDAQ standardized returns (solid) versus the normal distribution (dot-dash) over the period Feb 1971 to Feb 2001. Right: Corresponding quantile-quantile (QQ) plot with quantiles of the normal distribution on the abscissa and empirical quantiles on the ordinate. Returns are expressed as a %.

The major result of the APT then relates the expected premium of asset i to its exposure β_{ij} to factor j , and to each factor premium λ_j , $j = 1, \dots, k$. Specifically

$$\mathbb{E}X_i = r + \beta_{i1}\lambda_1 + \dots + \beta_{ik}\lambda_k, \quad (9)$$

where λ_j , $j = 1, \dots, k$, is the expected premium investors demand for bearing the risk of factor j . Notice that the factor premiums λ_j are the same for each security, and it is the exposure β_{ij} to each factor that depends on the security. Additionally if $k = 1$ in (8) and if we assume the existence of a risk-free asset r , $f_1 = X_M$ and that ϵ_i are uncorrelated with each other and the market, then $\lambda_1 = \mathbb{E}(X_M - r)$ and we get back the CAPM.

2.4 Empirical Evidence

Markowitz's mean-variance portfolio theory, as well as the CAPM and APT models, rely either explicitly or implicitly on the assumption of normally distributed asset returns⁸. Today, with long histories of price/return data available for a great many financial assets, it is easy to see that this assumption is inadequate. Empirical evidence suggests that asset returns have distributions which are heavier-tailed than the normal distribution. Figure 2 illustrates this for the *NASDAQ*⁹. The quantile-quantile (QQ) plot¹⁰ shows clearly that the distribution tails of the *NASDAQ* are heavier than the tails of the normal distribution. As early as 1963, Mandelbrot [Man63] and Fama [Fam65] rejected the assumption of

⁸As noted before, the multivariate normal assumption is consistent with maximizing expected utility.

⁹The daily *NASDAQ* time series, the corresponding returns and their maxima and minima are displayed in Figure 16. The time series starts in February 1971 and ends February 2001 (actually from February 08, 1971 to January 26, 2001). The corresponding empirical statistics can be found in Table 1.

¹⁰A quantile-quantile (QQ) plot is a graphical check to see if two distributions are of the same type. Two random variables X and Y are said to be of the same type if their distributions are the same up to a change in location and

Asset	Period	Mean	Std Dev	Skewness	Kurtosis ¹¹	Min	Max
<i>S&P 500</i>	01/51 - 03/2001	.033	.870	-1.61	43.9	-22.9	8.71
USD/GBP	02/1985 - 02/2001	.006	.677	.043	3.40	-4.13	4.59
TB/USD	02/85 - 03/2001	.011	.663	4.22	158	-8.57	17.8
<i>NASDAQ</i>	02/1971 - 02/2001	.044	1.08	-.523	15.5	-12.0	13.3

Table 1: Left: Empirical statistics for daily returns (as %) of several financial assets: the *S&P 500* index, the USD/British pound exchange rate, the Thai Baht/USD exchange rate and the *NASDAQ* composite index.

normality for other heavier-tailed distributions. In his 1963 paper, Mandelbrot not only confirmed the poor fit of the normal distribution, but proposed the model which is known today as the stable model for asset returns.

Recall that if the normal distribution is valid, then about 95% of the observations would lie within two standard deviations of the mean, and about 99% would lie within three standard deviations of the mean. In financial time series, large returns (both positive and negative) occur far too often to be compatible with the normal distribution assumption. The distribution of the financial return series are characterized not only by heavy tails, but also by a high peakedness at the center. In the Econometric terminology, they are said to be *leptokurtotic*.

To the risk manager trying to guard against large losses, the deviation from normality cannot be neglected. Suppose for example that daily returns are distributed as a Student- t distribution with 4 degrees of freedom (denoted t_4) and a variance given by σ^2 . Since this distribution has a much heavier tail than a normal distribution with the same variance, as one moves farther out into the tail of the distribution, rare events occur much more frequently. Figure 3 shows how much more likely *rare* events occur under the t_4 assumption than under the normal, when rare is defined in terms of standard deviations.

3 Value at Risk

In the early 1990s, a number of financial institutions (J.P. Morgan, Bankers Trust, ...) proposed a new risk measure to quantify by a single number the firms aggregate exposure to market risk. This measure, commonly known today as *value at risk* (VaR), is now used to measure not only market risk but other forms of risk to which the firm is exposed, such as credit, operational, liquidity, and legal risk. VaR is defined as the loss of a financial position over a time horizon τ that would be exceeded

scale. That is $X \stackrel{d}{=} aY + b$ for some $a \in \mathbb{R}^+$, $b \in \mathbb{R}$. Since the QQ plot plots quantiles of two distributions, if they are of the same type, the plot should be linear. In this case we are checking whether the empirical distribution of NASDAQ standardized returns and the hypothesized normal distribution are of the same type.

¹¹In this paper we use as definition of kurtosis

$$K(X) = \frac{\mathbb{E}(X - \mu_X)^4}{(\text{Var}X)^2} - 3,$$

so that the normal distribution has a kurtosis of zero. Heavy tails, therefore, will lead to positive kurtosis.

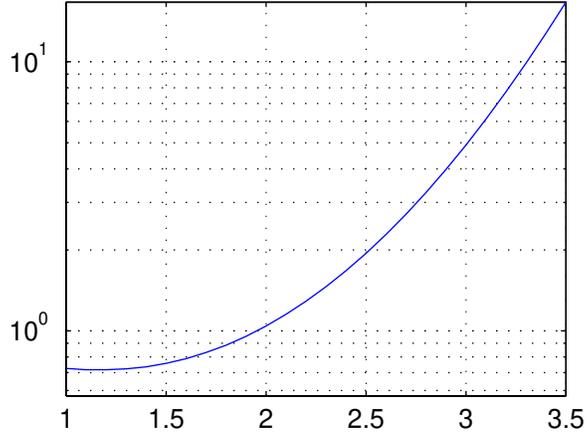


Figure 3: Ratio of tail probabilities $\mathbb{P}(T > k\sigma)/\mathbb{P}(X > k\sigma)$ plotted in units of k . Here $T \sim t_4$ and X is normal, both with variance σ^2 . T is more likely to take large values than X .

with small probability $1 - \alpha$, that is,

$$\mathbb{P}(\text{Loss} > \text{VaR}) \leq 1 - \alpha. \quad (10)$$

The confidence level α is typically a large number¹² between .95 and 1.

To define VaR precisely, let X be the random variable whose cumulative distribution function F_X describes the negative profit and loss distribution (P&L) of the risky financial position at the specified horizon time τ . Negative values of X correspond now to profits and positive values of X correspond to losses. This is a useful convention in risk management since there is then no ambiguity when discussing large losses (*large* values of X correspond to *large* losses).

Formally, value at risk is a *quantile* of the probability distribution F_X , that is roughly, the x corresponding to a given value of $0 < \alpha = F_X(x) < 1$.

Definition 3.1 *Let X be the random variable whose cumulative distribution function F_X describes the negative profit and loss distribution (P&L) of the risky financial position at the specified horizon time τ (so that losses are positive). Then, for a confidence level $0 < \alpha < 1$,*

$$\text{VaR}_\alpha(X) = \inf\{x | F_X(x) \geq \alpha\}. \quad (11)$$

We set, avoiding technicalities

$$\text{VaR}_\alpha(X) = F_X^{-1}(\alpha),$$

where F_X^{-1} denotes the inverse function of F_X ¹³ (see Figure 4). Hence the value $\text{VaR}_\alpha(X)$ over the horizon time τ would be exceeded on the average $100(1 - \alpha)$ times every 100τ time periods.

¹²In statistics, α and $1 - \alpha$ are usually interchanged because α , in statistics, denotes typically the Type 1 hypothesis testing error and is chosen small. The corresponding confidence level is then $1 - \alpha$.

¹³This is strictly correct when F_X is strictly increasing and continuous. Otherwise, one needs to use the *generalized*

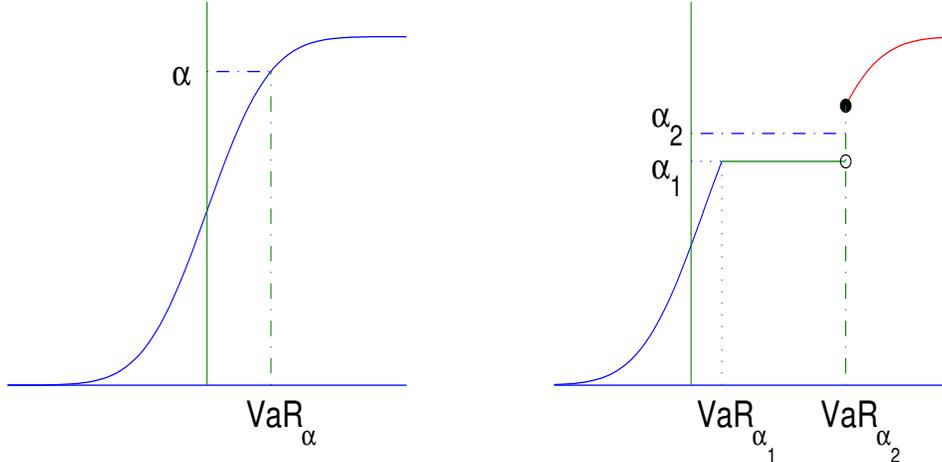


Figure 4: $VaR_\alpha(X)$ for different cumulative distributions functions (cdf) of the loss distribution X . The cdf on the right corresponds to an asset with discontinuous payoff, for example a binary option. See Definition 3.1.

Because of its intuitive appeal and simplicity, it is no surprise that VaR has become the *de facto* standard risk measure used around the world today. For example, today VaR is frequently used by regulators to determine minimum capital adequacy requirements. In 1995, the Basle Committee on Banking Supervision¹⁴ suggested that banks be allowed to use their own internal VaR models for the purpose of determining minimum capital reserves. The internal models approach of the Basle Committee is a ten day VaR at the $\alpha = 99\%$ confidence level multiplied by a *safety factor* of at least 3. Thus if $VaR = 1M$, the institution is required to have at least $3M$ in reserve in a safe account.

The safety factor of three is an effort by regulators to ensure the solvency of their institutions. It has also been argued, see Stahl [Sta97] or Danielsson, Hartmann and De Vries [DHV98], that the safety factor of three comes from the heavy-tailed nature of the return distribution. Since most VaR calculations are based on the simplifying assumption that the distribution of returns are normal¹⁵, how bad does this assumption effect VaR ? Assume that the Profit and Loss (P&L) distribution is symmetric and has finite variance σ^2 . Then regardless of the actual distribution, if X represents the random loss over the specified horizon time with mean zero, Chebyshev's inequality gives

$$\mathbb{P}[X > c\sigma] \leq \frac{1}{2c^2}.$$

So if we are interested in VaR bounds for $\alpha = 0.99$, setting $1/2c^2 = 0.01$ gives $c = 7.071$, and

inverse of F_X , denoted F_X^- , and defined as

$$F_X^-(\alpha) = \inf\{x \mid F_X(x) \geq \alpha\} \quad , \quad 0 < \alpha < 1.$$

The definition (11) of $VaR_\alpha(X)$ is then $VaR_\alpha(X) = F_X^-(\alpha)$. Thus, if $F_X(x) = \alpha$ for $x_0 \leq x \leq x_1$, then $VaR_\alpha(X) = F_X^-(\alpha) = x_0$.

¹⁴See [oBS5a] and [oBS5b]. Basle is a city in Switzerland. In French, Basle is Bâle, in German, it is Basel. Basle is the old name for the city. The accent in Bâle stands for the *s* that has been dropped from Basle.

¹⁵See for example the RiskMetrics manual [Met96].

this implies $VaR_{\alpha=.99}^{max}(X) = 7.071\sigma$. If the VaR calculation were done under the assumption of normality (Gaussian distribution) then $VaR_{\alpha=.99}^{Ga}(X) = 2.326\sigma$, and so if the true distribution is indeed heavy-tailed with finite variance then the correction for $VaR_{\alpha=.99}$ of three is reasonable, since $3 \times 2.326\sigma = 6.978\sigma$.

3.1 Computation of VaR

Before we discuss how $VaR_{\alpha}(X)$ is computed, we need to say a few words about X . Typically X represents the risk of some aggregated position which is influenced by many underlying risk factors Y_1, \dots, Y_d ,

$$X = f(Y_1, \dots, Y_d). \quad (12)$$

The functional form of the dependence of X on the factors Y_1, \dots, Y_d is usually never known exactly, but it may be approximated in several standard ways depending on the nature of the position. For example, f is linear in the case of a portfolio of straight equity positions. The function f is non-linear, for example, if the portfolio contains a call option on an equity since the value of the call changes non-linearly with respect to a change in the underlying asset. The usual procedure is to approximate the change in the calls value with respect to its underlying by the options delta. For small changes in the underlying such an approximation is reasonable. However for large changes in the underlying, the approximation can be quite bad. In an effort to improve the approximation, a second order term is sometimes added, the options gamma. This second order approximation is referred to as the delta-gamma approximation.

In practice, the VaR of a risky position X is calculated in one of three ways: through historical simulation, through a parametric model, or through some sort of Monte Carlo simulation. Each way involves assumptions and approximations and it is the responsibility of the user to be aware of them. The risk manager who blindly performs the model calculations does so at his or her peril. For a full treatment of the commonly used procedures for the calculation of VaR , see Jorion [Jor01], Dowd [Dow98] or Wilson [Wil98]. See Duffie and Pan [DP97] for a discussion of heavy tails and VaR calculations. We now describe the three ways of calculating VaR .

3.1.1 Historical Simulation VaR

The historical simulation model uses the historical returns of assets currently held in the portfolio in order to calculate VaR ¹⁶. First, returns over the horizon time τ are constructed for each asset in the portfolio using historical price information. Then portfolio returns are computed using the current weight distribution of assets as though the portfolio had been held during the whole historical period which is being sampled. The VaR is then read from the *historical sample* by using the order statistics. For example, if 1000 time periods are sampled, then 1000 portfolio returns are calculated, one for each time period. Let $X_p^{(1)} \geq X_p^{(2)} \geq \dots \geq X_p^{(1000)}$ be the order statistics of these returns, where losses are positive. Then $VaR_{\alpha=0.95}(X_p) = X_p^{(50)}$. The size of the sample is chosen by the user, but may be constrained by the available data for some of the assets currently held.

The model is simple to implement and has several advantages. Since it is based on historical prices it allows for a nonlinear dependence between assets in the portfolio and underlying risk factors. Also

¹⁶Over a fixed time horizon, VaR may be reported in units of rate of return (%) or of currency (profit and loss) since these are essentially the same, up to multiplication by the initial wealth/value.

since it uses historical returns it allows for the presence of heavy tails without making assumptions on the probability distributions of returns of the assets in the portfolio. There is therefore no *model risk*. In addition, there is no need to worry about the dependence structure of assets within the portfolio since it is already reflected in the price and return data.

The drawbacks are typical of models involving historical data. There may not be enough data available and there may be no reason to believe that the future will look like the past. For example, if the user would like to compute *VaR* for regulatory requirements, then $\tau = 10$ days. With about 260 business days, there are only 26 such observations in each year, four years worth of data are required to get about 100 historical simulations. This is the absolute minimum necessary to calculate *VaR* with $\alpha = .99$, since with 100 data points, there is but a single observation in the tail. If one or several of the assets in the portfolio have insufficient histories then adjustments must be made. For example, some practitioners bootstrap from the shorter return histories in order to take advantage of the longer histories on other assets.

When working only with historical data it is important to realize that we are assuming that the future will look like the past. If this assumption is likely to be unrealistic, the *VaR* estimate may be dangerously off the mark. For instance, if the sample period or window is devoid of large price changes, then our historical *VaR* will be low. But it will be large if there were large price fluctuations during the sample period. As large price fluctuations leave the sample window, the *VaR* will change accordingly. This yields a highly variable estimate and one which does not take into account the current financial climate. The deficiencies of historical simulation notwithstanding, its ease of use makes it the most popular method for *VaR* calculations.

3.1.2 Parametric *VaR*

The parametric *VaR* model assumes that the returns possess a specific distribution, usually normal. The parameters of the distribution are estimated using either historical data or forward looking option data.

Example 3.1 Assume that over the desired time horizon τ the (negative) return distribution of a portfolio is given by $F_X \sim \mathcal{N}(\mu_\tau, \sigma_\tau^2)$. Then the value at risk of portfolio X for horizon τ and confidence level $\alpha > 0.5$ is given by

$$\begin{aligned} VaR_\alpha(X) &= \inf\{x | F_X(x) \geq \alpha\} \\ &= F_X^{-1}(\alpha) \\ &= \mu_\tau + \sigma_\tau \Phi^{-1}(\alpha), \end{aligned}$$

where $\Phi^{-1}(\alpha)$ is the α quantile of the standard normal distribution.

More generally, if the (negative) return distribution of X is any F_X with finite mean μ_τ and finite variance σ_τ^2 , then

$$VaR_\alpha(X) = \mu_\tau + \sigma_\tau q_\alpha, \tag{13}$$

where q_α is the α quantile of the standardized version of X . In other words, $q_\alpha = F_{\tilde{X}}^{-1}(\alpha)$ where $\tilde{X} = (X - \mu_\tau)/\sigma_\tau$.

If the *VaR* is computed under the assumption that returns are light-tailed, say normal, when in fact they are heavy tailed, say t_ν (Student- t distribution with ν degrees of freedom), the risk may be

seriously underestimated for high confidence levels. This is because for large α , $F_{normal}^{-1}(\alpha) \leq F_{tv}^{-1}(\alpha)$, so that the value of x that achieves $F_{normal}(x) = \alpha$ is smaller than the value of x that achieves $F_{tv}(x) = \alpha$. It is thus very important that the return distribution be modelled well. A wide variety of parametric distributions can be considered.

Within the portfolio context, the most easily implemented parametric model is the so called *delta-normal* method, where the joint distribution of the risk factor returns is multivariate normal and the returns of the portfolio are assumed to be a linear function of the returns of the underlying risk factors. In this case the portfolio returns are themselves normally distributed.

Example 3.2 Take a portfolio of equities whose (negative) returns are given by $X_p = w_1X_1 + \dots + w_nX_n$ where w_i is the weight given to asset i and X_i is the assets (negative) return over the horizon in question. Assume $(X_1, \dots, X_n) \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Then, for $\alpha \in (0.5, 1)$,

$$\begin{aligned} VaR_\alpha(X_p) &= \Phi^{-1}(\alpha)\sqrt{\mathbf{w}^T \Sigma \mathbf{w}} \\ &= \sqrt{\overrightarrow{VaR}_\alpha^T \boldsymbol{\rho} \overrightarrow{VaR}_\alpha}, \end{aligned}$$

where $\overrightarrow{VaR}_\alpha = (VaR_\alpha(w_1X_1), \dots, VaR_\alpha(w_nX_n))$ is the vector of the individual weighted asset *VaRs* and $\boldsymbol{\rho}$ is the asset return correlation matrix. See Dowd [Dow98] for details.

When the number of assets is large, the central limit theorem is often invoked in defense of the normal model. Even if the individual asset returns are non-normal, the central limit theorem tells us that the weighted sum of many assets should be approximately normal. This argument may be disposed of in various ways. Consider, for example, the empirical distribution of daily returns of a large diversified index such as the *NASDAQ*, which is clearly heavy-tailed (see Figure 2). From a probabilistic point of view it is not at all obvious that the assumptions of the central limit theorem are satisfied. For example, if the returns do not have finite variance, there may be convergence to the class of stable distributions.

The class of stable distributions (also known as α -stable or *stable Paretian*) may be defined in a variety of ways. More will be said about them in Section 7. We define, at this stage, a stable distribution as the only possible limiting distribution of appropriately normalized sums of independent random variables.

Definition 3.2 *The random variable X has a stable distribution if there exists a sequences of i.i.d. random variables $\{Y_i\}$ and constants $\{a_n\} \in \mathbb{R}$ and $\{b_n\} \in \mathbb{R}^+$ such that*

$$\frac{Y_1 + \dots + Y_n}{b_n} - a_n \xrightarrow{d} X \quad \text{as } n \rightarrow \infty. \quad (14)$$

The stable distribution of X in (14) is characterized by four parameters $(\alpha, \sigma, \beta, \mu)$ and we write $X \sim S_\alpha(\beta, \sigma, \mu)$. The parameter $\alpha \in (0, 2]$ is called the *index of stability* or the *tail exponent* and controls the decay in the tails of the distribution. The remaining parameters σ, β, μ control scale, skewness, and location respectively. If the Y_i have finite variance (the case in the usual CLT) then $\alpha = 2$ and the distribution of X is Gaussian. For all $\alpha \in (0, 2)$ the distribution is non-Gaussian stable and possess heavy tails.

Example 3.3 Properties of weekly returns of the Nikkei 225 Index over a 12 year period are examined in Mittnik, Rachev and Paolella [MRP98]. The authors fit the return distribution using a number of parametric distributions, including the normal, Student- t and stable. According to various measures of goodness of fit, the partially asymmetric Weibull, Student- t and the asymmetric stable provide the best fit. The fit by the normal is shown to be relatively poor. The stable distribution, in addition, fits best the tail quantiles of the empirical distribution, which is a result most relevant to the calculation of VaR .

The central limit theorem typically assumes independence. Although it has extensions to allow for mild dependence, this dependence must be sufficiently weak. In fact, for a given number of assets, the greater the dependence, the worse the normal approximation. This affects the speed of the convergence. Since a VaR calculation involves the tails of the distribution, it is most important that the approximation hold in the tails. However, even when the conditions for the central limit theorem hold, the convergence in the tail is known to be very slow. The normal approximation may then only be valid in the central part of the distribution. In this case, the return distribution may be better approximated by a heavier-tailed distribution such as the Student- t or *hyperbolic* whose use in finance is becoming more common.

The hyperbolic distribution is a subclass of the class of generalized hyperbolic distributions. The generalized hyperbolic distributions were introduced in 1977 by Barndorff-Neilsen [BN77] in order to explain empirical findings in geology. Today these distributions are becoming popular in finance, and in particular in risk management. Two subclasses, the hyperbolic and the *inverse Gaussian*, are most commonly used. Both these subclasses may be shown to be mixtures of Gaussians. As such, they possess heavier tails than the normal distribution but not as heavy as the stable distribution. For an introduction to generalized hyperbolic distributions in finance, see for example Eberlein and Keller [EK95], Eberlein and Prause [EP00] or Shiryaev [Shi99].

3.1.3 Monte Carlo VaR

Monte Carlo procedures are perhaps the most flexible methods for computing VaR . The risk manager specifies a model for the underlying risk factors, which incorporates somehow their dependence. For example, the risk factors in (12) may be described by the *stochastic differential equation*

$$dY_t^{(i)} = Y_t^{(i)}(\mu_t^{(i)} dt + \sigma_t^{(i)} dW_t^{(i)}), \quad (15)$$

for $i = 1, \dots, d$, where $\mathbf{W}_t = (W_t^{(1)}, \dots, W_t^{(d)})$ is a multivariate Wiener process. Once parameters of the model are estimated, for example by using historical data, or option implied estimates, the risk factors paths are then computer generated, thousands of paths for each risk factor. Each set of simulated paths for the risk factors yields a portfolio path and the portfolio is priced accordingly. Each computed price of the portfolio represents a point on the portfolio's return distribution. After many such points are obtained the portfolio's VaR may then be read off the simulated distribution.

This method has the advantage of being extremely versatile. It allows for heavy tails, non-linear payoffs and a great many other user specifications. Within the Monte Carlo framework, risk managers may use their own pricing models to determine non-linear payoffs under many different scenarios for the underlying risk factors. The method has also the advantage of allowing for time varying parameters within the risk factor processes. See for example Broadie and Glasserman [BG98].

There are two major drawbacks to Monte Carlo methods. First, they are computationally very expensive. Thousands of simulations of the risk factors may have to be carried out for results to be trusted. For a portfolio with a large number of assets this procedure may quickly become unmanageable, since each asset within the portfolio must be valued using these simulations. Second, the method is prone to model risk. The risk factors and the pricing models of assets with non-linear payoffs may both be mis-specified. And, as is the case of the parametric *VaR*, there is the risk of mis-specifying the model parameters.

3.2 Parameter Estimation

The parametric and Monte Carlo *VaR* methods require parameters to be estimated. When one is interested in short time horizons, the primary goal is to estimate the volatility and covariance/correlation¹⁷. We outline some of the common estimation techniques here.

3.2.1 Historical Volatility

There are two different approaches to modelling volatility and covariance using only historical data. The more common approach gives constant weights to each data point. It assumes that volatility and covariance are constant over time. The other approach attempts to address the fact that volatility and covariance are time dependent by giving more weight to the more recent data points in the sample window.

First assume that variances and covariances do not to change over time. Take a large window of length n in which historical data on the risk factors is available. Let Y_{i,t_k} be the return of factor i at time period t_k . The variance of factor i and covariance of factors i and j are then computed by giving equal weights to each data point in the past. The n -period estimates at time T for the variance and covariance

$$\hat{\sigma}_i^2 = \frac{1}{n-1} \sum_{t=T-n}^{T-1} (Y_{i,t} - \hat{\mu}_{Y_i})^2 \quad \text{where} \quad \hat{\mu}_{Y_i} = \frac{1}{n} \sum_{t=T-n}^{T-1} Y_{i,t} \quad (16)$$

and

$$\hat{\sigma}_{i,j} = \frac{1}{n-1} \sum_{t=T-n}^{T-1} (Y_{i,t} - \hat{\mu}_{Y_i})(Y_{j,t} - \hat{\mu}_{Y_j}) \quad (17)$$

respectively¹⁸. Since equal weight is given to each data point in the sample, the estimated volatility and covariance change only slowly. If one keeps the window length fixed, the estimated values will rise or fall as new large returns enter the sample period and old large returns leave it. This means that even a single extreme return will affect the estimates in the same way, whether it occurred at time $T-1$ or time $T-n$. The estimated variance and covariance, therefore, are greatly influenced by the choice of the window size n .

Another *stylized fact* of financial time series, however, is that volatility itself is volatile. With this in mind, another historical estimate of variance and covariance uses a weighting scheme which gives

¹⁷For example, over short time horizons, the mean return is usually assumed to be zero.

¹⁸The normalization constant $n-1$ gives an unbiased estimate. It is sometimes replaced by n in order to correspond to the maximum likelihood estimate.

more weight to more recent observations. The corresponding estimates of variance and covariance are

$$\begin{aligned}\hat{\sigma}_i^2(T) &= \sum_{t=T-n}^{T-1} \alpha_t (Y_{i,t} - \hat{\mu}_{Y_i})^2, \\ \hat{\sigma}_{i,j}(T) &= \sum_{t=T-n}^{T-1} \alpha_t (Y_{i,t} - \hat{\mu}_{Y_i})(Y_{j,t} - \hat{\mu}_{Y_j}),\end{aligned}$$

where the weights α_t , $\sum_{t=T-n}^{T-1} \alpha_t = 1$, are chosen to reflect current volatility conditions. In particular, more weight is given to recent observations: $1 > \alpha_{T-1} > \alpha_{T-2} > \dots > \alpha_{T-n} > 0$. The model using exponentially decreasing weights, such as that used by RiskMetrics, is probably the most popular. In RiskMetrics, the volatility estimator is given by

$$\hat{\sigma}_i(T) = \sqrt{(1-\lambda) \sum_{t=1}^n \lambda^{t-1} (Y_{i,T-t} - \hat{\mu}_{Y_i})^2} \quad (18)$$

where the decay factor λ is chosen to best match a large group of assets¹⁹. The covariance estimate is similar. RiskMetrics choses $\lambda = 0.94$ in the case of daily returns.

The choice (18) allows the forecast of the next periods volatility given the current information, and hence to make parametric *VaR* calculations given the current information. To see this, assume that the time T (negative) return distribution X_T is being modelled by

$$X_T \stackrel{d}{=} \sigma_T Z_T \quad (19)$$

where $Z_t, t \in \mathbb{Z}$, is an innovation process, that is a sequence of *i.i.d.* mean zero and unit variance random variables. Letting \mathcal{F}_t denote the filtration²⁰ we have

$$\begin{aligned}\sigma_{T+1|\mathcal{F}_T}^2 &= (1-\lambda) \sum_{t=0}^{\infty} \lambda^t X_{i,T-t}^2 \\ &= (1-\lambda) X_T^2 + \lambda(1-\lambda)(X_{T-1}^2 + \lambda X_{T-2}^2 + \lambda^2 X_{T-3}^2 + \dots) \\ &= (1-\lambda) X_T^2 + \lambda \sigma_{T|\mathcal{F}_{T-1}}^2.\end{aligned}$$

This allows us to make our *VaR* calculation depend on the conditional return distribution $F_{X_{T+1}|\mathcal{F}_T}$. If $VaR_\alpha^{T+1}(X)$ denotes the estimated value at risk for X at confidence level α for the period $T+1$ at time T , then, by (19),

$$VaR_\alpha^{T+1}(X) = \sigma_{T+1|\mathcal{F}_T} q_\alpha,$$

¹⁹In this estimate it is assumed that the decay parameter λ and window length n are such that the approximation

$$\sum_{t=1}^n \lambda^{t-1} \cong \frac{1}{1-\lambda}$$

is valid.

²⁰Conditioning over \mathcal{F}_T means conditioning over all the observations X_1, \dots, X_T .

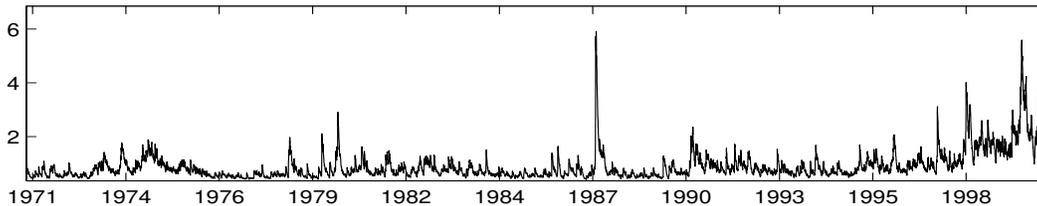


Figure 5: GARCH(1,1) volatilities σ_t for NASDAQ.

where q_α is the α quantile of the innovation process Z_{t+1} . In RiskMetrics Z is $\mathcal{N}(0, 1)$, in which case the return process X_t is *conditionally* normal²¹.

The modelling of the volatility using exponential weights and the assumption of conditional normality has two major effects. First, the volatility estimator, which is now truly time varying, attempts to account for the local volatility conditions by giving more weight to the most recent observations. It also has a second less obvious, but no less profound effect on the calculation of VaR . Even though the conditional return distribution may be assumed to be normal (thin-tailed) within the VaR calculation, the unconditional return distribution will typically have heavier tails than the normal. This result is not surprising since we may think of our time t return as being sampled from a normal distribution with changing variance. This means that our unconditional distribution is more likely to fit the empirical returns and thus to provide a better estimate of the true VaR .

3.2.2 ARCH/GARCH Volatilities

The ARCH/GARCH class of conditional volatility models were first proposed by Engle [Eng82] and Bollerslev [Bol86] respectively. We will again assume that the (negative) return process to be modelled is of the form (19) where Z_t are *i.i.d.* mean zero, unit variance random variables representing the innovations of the return process. In the GARCH(p, q) model²², the conditional variance is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

In its most common form, $Z_t \sim \mathcal{N}(0, 1)$, so that the returns are conditionally normal. Just as in the exponentially weighted model for volatility (see Section 3.1.1), the GARCH model with a conditionally normal return distribution can lead to heavy tails in the unconditional return distribution. In the case of the GARCH(1, 1) model

$$\begin{aligned} X_t &= \sigma_t Z_t \quad \text{where } Z_t \sim \mathcal{N}(0, 1) \quad i.i.d., \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

²¹RiskMetrics allows the assumption of conditional normality to be relaxed in favor of heavier-tailed conditional distributions. For example the conditional distribution of returns may be mixture of normals or a generalized error distribution, that is, a double sided exponential.

²²The ARCH(p) model first proposed by Engle is equivalent to the GARCH(p, 0) model later proposed by Bollerslev. The advantage of the GARCH model over the ARCH model is that it requires fewer parameters to be estimated, because AR models (ARCH) of high order are often less parsimonious than ARMA models (GARCH) of lower order.

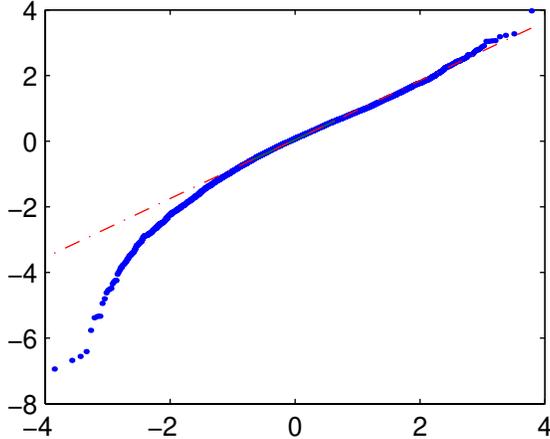


Figure 6: Quantile-quantile (QQ) plot of the conditionally normal GARCH(1,1) standardized *ex post* innovations for *NASDAQ* with the $\mathcal{N}(0,1)$ distribution.

it is straightforward to show that under certain conditions²³ the unconditional centered kurtosis is given by

$$K = \frac{\mathbb{E}X_t^4}{(\mathbb{E}X_t^2)^2} - 3 = \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2},$$

which for most financial return series will be greater than zero. For example, in the case of a stationary ARCH(1) model, $X_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2} Z_t$, with $\alpha_0 > 0$ and $\alpha_1 \in (0, 2e^\gamma)$, where γ is Euler's constant²⁴, Embrechts, Klüppelberg and Mikosch [EKM97] show that the unconditional distribution is formally heavy-tailed, that is

$$\mathbb{P}(X > x) \sim cx^{-\alpha}, \quad x \rightarrow \infty. \quad (20)$$

where $\alpha/2 > 0$ is the unique solution to the equation $h(u) = \frac{(2\alpha_1)^u}{\sqrt{\pi}} \Gamma(u + \frac{1}{2}) = 1$.

The ARCH/GARCH models allow for both volatility clustering (periods of large volatility) and for heavy tails. The GARCH(1,1) estimated volatility process σ_t for the *NASDAQ* is displayed in Figure 5. The assumption of conditional normality can be checked, for example, by examining a QQ plot of the *ex post* innovations, that is $\hat{Z}_t = X_t/\hat{\sigma}_t$. Figure 6 displays the QQ plot of \hat{Z}_t in the traditional, conditionally normal GARCH(1,1) model for the *NASDAQ*. The fit of the GARCH(1,1) conditionally normal model in the lower tail is poor, showing the lower tail of \hat{Z}_t is heavier than the normal distribution.

If the distribution of the historical innovations Z_{t-n}, \dots, Z_t is heavier-tailed than the normal, one can modify the model to allow a heavy-tailed conditional distribution $F_{X_{t+1}|\mathcal{F}_t}$ ²⁵. In Panorska,

²³These conditions are $\alpha_1 + \beta_1 < 1$ to guarantee stationarity, and $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ for $K > 0$. Both are generally met in financial time series.

²⁴Euler's constant γ is given by $\gamma = \lim_{n \rightarrow \infty} (\sum_{k=1}^n \frac{1}{k} - \ln n)$ and is approximately $\gamma \approx 0.577$.

²⁵For example the GARCH module in the statistical software package SPlus allows for three different non-Gaussian conditional distributions. As long as the user can estimate the GARCH parameters, usually through maximum likelihood, there are virtually no limits to the choice of the conditional distribution.

Mittnik and Rachev [PMR95] and Mittnik, Paoletta and Rachev [MPR97], returns on the Nikkei index are modelled using an ARMA-GARCH model of the form

$$X_t = a_0 + \sum_{i=1}^r a_i X_{t-i} + \epsilon_t + \sum_{j=1}^s b_j \epsilon_{t-j} \quad (21)$$

(contrast with (19)), where $\epsilon_t = \sigma_t Z_t$, with Z_t an *i.i.d.* location zero, unit scale heavy-tailed random variable. The conditional distribution of the return series $F_{X_t|\mathcal{F}_{t-1}}$ is given by the distribution type of Z_t . The ARMA structure in (21) is used to model the conditional mean $\mathbb{E}(X_t|\mathcal{F}_{t-1})$ of the return series X_t . The GARCH structure is imposed on the scale parameter²⁶ σ_t through

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

Several choices for the distribution of Z_t are tested. In the case where Z_t are realizations from a stable distribution, the GARCH model used is

$$\sigma_t = \alpha_0 + \sum_{i=1}^p \alpha_i |\epsilon_{t-i}| + \sum_{j=1}^q \beta_j \sigma_{t-j},$$

and the index of stability exponent α for the stable distribution is constrained to be greater than one.

Using several goodness of fit measures, the authors find that it is better to model the conditional distribution of returns for the Nikkei than the unconditional distribution, since the unconditional distribution cannot capture the observed temporal dependencies of the return series²⁷. Within the tested models for Z_t , the partially asymmetric Weibull, the Student- t , and the asymmetric stable all outperform the normal. In order to perform reliable value at risk calculations one must model the tail of the distribution Z_t particularly well. The Anderson-Darling (AD) statistic can be used to measure goodness of fit in the tails. Letting $F_{emp}(x)$ and $F_{hyp}(x)$ denote the empirical and hypothesized parametric distributions respectively, the AD statistic

$$AD = \sup_{x \in \mathbb{R}} \frac{|F_{emp}(x) - F_{hyp}(x)|}{\sqrt{F_{hyp}(x)(1 - F_{hyp}(x))}}$$

gives more weight to the tails of the distribution. Using this statistic, as well as others, the authors propose the asymmetric stable distribution as the best of the tested models for performing *VaR* calculations at high quantiles.

The class of ARCH/GARCH models have become increasingly popular for computing *VaR*. The modelling of the conditional distribution has two immediate benefits. First, it allows for the predicted volatility (or scaling) to use local information, *i.e.* it allows for volatility clustering. Second, since volatility is allowed to be volatile, the unconditional distribution will typically not be thin-tailed. This is true, as we have seen, even when the conditional distribution is normal.

There now exist many generalizations of the class of ARCH/GARCH models. Models such as EGARCH, HGARCH, AGARCH, and others, all attempt to use the local volatility structure to better

²⁶In their model σ_t is to be interpreted as a scale parameter, not necessarily a volatility, since for some of the distributional choices for Z_t , the variance may not exist.

²⁷The type of the conditional distribution is that of Z_t , the unconditional distribution is that of X_t .

predict future volatility while trying to account for other observed phenomenon. See Bollerslev, Chou and Kroner [BCK92] for a review. The time series of returns $\{X_t\}_{t \in \mathbb{Z}}$ in (19) is generally assumed to be stationary. In a recent paper, Mikosch and Stărică [MS00] show that this assumption is not supported, at least globally, by the *S&P 500* from 1953 to 1990 and the DEM/USD foreign exchange rate from 1975 to 1982. The authors show that when using a GARCH model the parameters must be updated to account for changes of structure (changes in the unconditional variance) of the time series. A method for detecting these changes is also proposed. Additionally, they show that the long range dependence behavior associated with the absolute return series, another of the so called *stylized facts* of financial time series, may only be an artifact of structural changes in the series, that is, to non-stationarity.

Stochastic volatility models are not limited to the class of ARCH/GARCH models and their generalizations. Other models may involve additional sources of randomness. For example, the model of Hull and White [HW87]

$$\begin{aligned} dY_t &= \mu Y_t + \sigma_t Y_t dW_t^{(1)}, \\ dV_t &= \nu V_t + \xi V_t dW_t^{(2)}, \end{aligned}$$

where $\sigma_t^2 = V_t$ and $(W_t^{(1)}, W_t^{(2)})$ is a bivariate Wiener process, introduces a second source of randomness through the volatility. The two sources of randomness $W_t^{(1)}$ and $W_t^{(2)}$ need not be uncorrelated. Again, the introduction of a stochastic scaling generally leads to an unconditional return distribution which is leptokurtotic. See Shiryaev [Shi99], for an introduction to stochastic volatility models in discrete and continuous time.

3.2.3 Implied Volatilities

The parametric *VaR* calculation requires a forecast of the volatility. All of the models examined so far have used historical data. One may prefer to use a *forward* looking data set instead of historical data in the forecast of volatility, for example options data, which provide the market estimate of future volatility. To do so, one could use the *implied volatility* derived from the Black-Scholes model. In this model, European call options prices $C_t = C(S_t, K, r, \sigma, T - t)$ are an increasing function of the volatility σ . The stock price S_t at time t , the strike price K , the interest rate r and the time to expiration $T - t$ are known at time t . Since σ is the only unknown parameter/variable, we may then use the observed market price C_t to solve for σ . This estimate of σ is commonly called the (Black-Scholes) implied volatility. The Black-Scholes model, however is imperfect. While σ should be constant, one typically observes that σ depends on the time to expiration $T - t$ and on the strike price K . For fixed $T - t$, the implied volatility $\sigma = \sigma(T - t, K)$ as a function of the strike price K is often convex, a phenomenon known as the *volatility smile*. To obtain volatility estimates it is common to use at-the-money options, where $S_t = K$, since they are the most actively traded and hence are thought to provide the most accurate estimates.

3.2.4 Extreme Value Theory

Since *VaR* calculations are only concerned with the tails of a probability distribution, techniques from Extreme Value Theory (EVT) may be particularly effective. Proponents of EVT have made compelling arguments for its use in calculating *VaR* and for risk management in general. We will discuss EVT in Section 6.

4 Risk Measures

We have considered two different measures of risk: standard deviation and value at risk. Standard deviation, used by Markowitz and others, is still commonly used in portfolio theory today. The second measure, VaR , is the standard measure used today by regulators and investment banks. We detailed some of the computational issues surrounding these measures but have not discussed their validity.

It is easy to criticize standard deviation and value at risk. Even in Markowitz's pioneering work on portfolio theory, the shortcomings of standard deviation as a risk measure were recognized. In [Mar59], an entire chapter is devoted to *semi-variance*²⁸ as a potential alternative. In Artzner, Delbaen, Eber and Heath [ADEH97], for example, measures based on standard deviation are criticized based on their inability to describe rare events and VaR is criticized because of its inability to aggregate risks in a logical manner. In two now famous papers [ADEH97] and [ADEH99] on financial risk, the authors propose a set of properties any reasonable risk measure should satisfy. Any risk measure which satisfies these properties is called *coherent*. We shall now introduce these properties and indicate why the risk measures described above are not coherent.

4.1 Coherent Risk Measures

Suppose that the financial position of an investor will lead at time T to a loss X ²⁹, which is a random variable. Let \mathcal{G} be the set of all such X . A risk measure ρ is defined as a mapping from \mathcal{G} to \mathbb{R} . Intuitively, for a given potential loss X in the future we may think of $\rho(X)$ as the minimum amount of cash that we need to invest prudently today (in a reference instrument) to be allowed to take the position X ³⁰. A risk measure ρ may be *coherent* or not.

Definition 4.1 *Given a reference instrument with return r , possibly random, a risk measure ρ satisfying the following four axioms is said to be coherent:*

Translation Invariance. For all $X \in \mathcal{G}$ and all $\alpha \in \mathbb{R}$, we have $\rho(X + \alpha r) = \rho(X) + \alpha$. This means that adding the amount α to the position, and investing it prudently, reduces the overall risk of the position by α .

Subadditivity. For all X_1 and $X_2 \in \mathcal{G}$, $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$. Hence a merger does not create extra risk. This is the basis for diversification.

Positive Homogeneity. For all $\lambda \geq 0$ and all $X \in \mathcal{G}$, $\rho(\lambda X) = \lambda \rho(X)$. This requires that the risk scales with the size of a position. If the size of a position renders it illiquid, then this should be considered when modelling the future net worth.

Monotonicity. For all X and $Y \in \mathcal{G}$ with $X \geq Y$, we have $\rho(X) \geq \rho(Y)$. If the future net loss X is greater, then X is more risky.

²⁸In order to put the accent on (negative) returns above the mean, semi-variance is defined as

$$\tilde{\sigma}_X = \mathbb{E}[(X - \mathbb{E}X)1_{\{X > \mathbb{E}X\}}]^2.$$

²⁹Losses are positive and profits negative. This is at odds with the authors' original notation.

³⁰The authors refer to X as *risk* and axiomatically define *acceptance sets*, which are sets of acceptable risks, and proceed to define measures of risk as describing the risks proximity to the acceptance set.

The term *coherent measure of risk* has found its way into the risk management vernacular. It is defined, for example, in the second edition of Philippe Jorion’s *Value at Risk* ([Jor01]).

Note that the axioms of translation invariance and monotonicity rule out standard deviation as a coherent measure of risk. Indeed, since $\sigma_{X+\alpha r} = \sigma_X$, translation invariance fails, and since σ also penalizes the investor for large profits as well as large losses, monotonicity fails as well. Consider, for example, two portfolios X and Y which are identical except for the free lottery ticket held in Y . We have $X \geq Y$, since there is no down-side to the free ticket and therefore the potential losses in Y are smaller than in X . Nevertheless, the standard deviation measure assigns to Y a higher risk, hence monotonicity fails. Markowitz’s alternative risk measure semi-variance is not coherent either because it is not subadditive.

4.2 Expected Shortfall

VaR is not a coherent measure of risk because it fails to be subadditive in general. One can indeed easily construct scenarios (see Albanese [Alb97]) where for two positions X and Y it is true that

$$VaR_\alpha(X + Y) > VaR_\alpha(X) + VaR_\alpha(Y).$$

This is contrary to the risk managers feelings, that the overall risk of different trading desks is bounded by the sum of their individual risks. In short, VaR fails to aggregate risks in a logical manner. In addition, VaR tells us nothing about the size of the loss that exceeds it. Two distributions may have the same VaR yet be dramatically different in the tail.

Hence neither the standard deviation nor VaR are coherent. On the other hand, the *expected shortfall*, also called *tail conditional expectation*, is a coherent risk measure. Intuitively, the expected shortfall addresses the question: *given that we will have a bad day, how bad do we expect it to be?* It is a more conservative measure than VaR and looks at the average of all losses that exceed VaR . Formally, the expected shortfall for risk X and high confidence level α is defined as follows:

Definition 4.2 *Let X be the random variable whose distribution function F_X describes the negative profit and loss distribution (P&L) of the risky financial position at the specified horizon time τ (thus losses are positive). Then the expected shortfall for X is*

$$S_\alpha(X) = \mathbb{E}(X|X > VaR_\alpha(X)). \tag{22}$$

Suppose, for example, that a portfolio’s risk is to be calculated through simulation. If 1000 simulations are run, then for $\alpha = 0.95$, the portfolios VaR would be the smallest of the 50 largest losses. The corresponding expected shortfall would be estimated by the numerical average of these 50 largest losses. Expected shortfall, therefore, tells us something about the expected size of a loss exceeding VaR . It is subadditive, coherent and puts fewer restrictions on the distribution of X , requiring only a finite first moment to be well defined. Additionally, it may be reconciled with the idea of maximizing expected utility. Levy and Kroll [LK78] show that for all utility functions U with the properties described in Section 2.1 and all random variables X and Y (representing losses) that

$$\mathbb{E}U(-X) \geq \mathbb{E}U(-Y) \iff S_\alpha(X) \leq S_\alpha(Y) \quad \text{for all } \alpha \in (0, 1).$$

Expected shortfall can be used in portfolio theory as a replacement of the standard deviation if the distribution of \mathbf{X} is normal, or more generally, elliptical. As we will see in Section 5.3, in this

case any positive homogeneous translation invariant risk measure will yield the same optimal linear portfolio for the same level of expected return.

Unlike standard deviation, expected shortfall, as defined in (22), does not measure deviation from the mean. Bertsimas, Lauprete and Samarov [BLS00] define shortfall³¹ as

$$s_\alpha(X) = \mathbb{E}(X \mid X > VaR_\alpha(X)) - \mathbb{E}X. \quad (23)$$

The subtraction of the mean makes it more similar to the standard deviation $\sigma_X = \sqrt{\mathbb{E}(X - \mathbb{E}X)^2}$ and again, as far as portfolio theory is concerned, in the case of elliptical distributions, one obtains the same optimal portfolio for the same level of expected return if one uses s_α to measure risk. In fact, it can be shown that for a linear portfolio $X_p = w_1X_1 + \dots + w_nX_n$ of multivariate normally distributed returns $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, that

$$s_\alpha(X_p) = \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha} \sigma_p,$$

where $\phi(x)$ and $\Phi(x)$ are respectively, the pdf and cdf of a standard normal random variable evaluated at x . In other words,

$$\arg \min_{\mathbf{A}\mathbf{w}=\mathbf{b}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} = \arg \min_{\mathbf{A}\mathbf{w}=\mathbf{b}} s_\alpha(\mathbf{w}^T \mathbf{X}),$$

for all $\alpha \in (0, 1)$, where $\mathbf{A}\mathbf{w} = \mathbf{b}$ is *any* set of linear constraints, including constraints that do not require all portfolios to have the same mean. Note, however, that s_α is not coherent since it violates the axioms of translation invariance and monotonicity.

5 Portfolios and Dependence

The measure of dependence most popular in the financial community is linear correlation³². Its popularity may be traced back to Markowitz' mean variance portfolio theory since, under the assumption of multivariate normality, the correlation is the canonical measure of dependence. Outside of the world of multivariate normal distributions, correlation as a measure of dependence may lead to misleading conclusions (see Section 5.2.1)³³. The linear correlation between two random variables X and Y , defined by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (24)$$

is a measure of *linear dependence* between X and Y . The word *linear* is used because when variances are finite, $\rho(X, Y) = \pm 1$ if and only if Y is an affine transformation of X *almost surely*, that is if $Y = aX + b$ a.s. for some constants $a \in \mathbb{R} \setminus \{0\}$, and $b \in \mathbb{R}$. When the distribution of returns \mathbf{X} is multivariate normal, the dependence structure of the returns is determined completely by the covariance matrix $\boldsymbol{\Sigma}$ or, equivalently, by the correlation matrix $\boldsymbol{\rho}$. One has $\boldsymbol{\Sigma} = [\boldsymbol{\sigma}] \boldsymbol{\rho} [\boldsymbol{\sigma}]$ where $[\boldsymbol{\sigma}]$ is a diagonal matrix with the standard deviations σ_j on the diagonal.

When returns are not multivariate normal, linear correlation may no longer be a meaningful measure of dependence. To deal with potential alternatives, we will introduce the concept of copulas, describe various measures of dependence and focus on elliptical distributions. For additional details

³¹We still assume losses are positive. This is at odds with the authors notation.

³²Also known as Pearson's correlation.

³³Linear correlation is actually the canonical measure of dependence for the class of elliptical distributions. This class will be introduced shortly and may be thought of as an extension of multivariate normal distributions.

and proofs, see Embrechts, McNeil and Straumann [EMS01], Lindskog [Lin00b], Nelsen [Nel99], Joe [Joe97] and Fang, Kotz and Ng [FKN90].

5.1 Copulas

When $\mathbf{X} = (X_1, \dots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the distribution of any linear portfolio of the X_j 's is normal with known mean and variance. In the non-normal case, the joint distribution of \mathbf{X} ,

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

is not fully described by its mean and covariance. One would like, however, to describe the joint distribution by specifying separately the marginal distributions, that is, the distribution of the components X_1, \dots, X_n , and the dependence structure. One can do this with copulas.

Definition 5.1 *An n -Copula is any function $C : [0, 1]^n \rightarrow [0, 1]$ satisfying the following properties:*

1. *For every $\mathbf{u} = (u_1, \dots, u_n)$ in $[0, 1]^n$ we have that $C(\mathbf{u}) = 0$ if at least one component $u_j = 0$ and $C(\mathbf{u}) = u_j$ if $\mathbf{u} = (1, \dots, 1, u_j, 1, \dots, 1)$.*
2. *For every $\mathbf{a}, \mathbf{b} \in [0, 1]^n$ such that $\mathbf{a} \leq \mathbf{b}$*

$$\sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1 + \dots + i_n} C(u_{1i_1}, \dots, u_{ni_n}) \geq 0 \quad (25)$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for $j = 1, \dots, n$.

Corollary 5.1 below provides a concrete way to construct copulas. It is based on the following theorem due to Sklar (see [Skl96], [Nel99]), which states that by using copulas one can separate the dependence structure of the multivariate distribution from the marginal behavior.

Theorem 5.1 (Sklar) *Let F be an n -dimensional distribution function with marginals $X_j \sim F_j$ for $j = 1, \dots, n$. Then there exists an n -copula $C : [0, 1]^n \rightarrow [0, 1]$ such that for every $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$,*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)). \quad (26)$$

Furthermore, if the F_j are continuous then C is unique. Conversely, if C is an n -copula and F_j are distribution functions, then F in (26) is an n -dimensional distribution function with marginals F_j .

The function C is called the copula of the multivariate distribution of \mathbf{X} . Assuming continuity of the marginals F_j , $j = 1, \dots, n$, we see that the copula C of F is the joint distribution of the uniform transformed variables $F_j(X_j)$,

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)). \quad (27)$$

Corollary 5.1 *If the F_j are the cdfs of $U(0, 1)$ random variables, then $x_j = F_j(x_j)$, $0 < x_j < 1$, and (26) becomes $F(x_1, \dots, x_n) = C(x_1, \dots, x_n)$. Therefore the copula C may be thought of as the cumulative distribution function (cdf) of a random vector with uniform marginals.*

Copulas allow us to model the joint distribution of \mathbf{X} in two natural steps. First, one models the univariate marginals X_j . Second, one chooses a copula that characterizes the dependence structure of the joint distribution. Any n -dimensional distribution function can serve as a copula. The following examples relate familiar multivariate distributions to their associated copulas and marginals.

Example 5.1 Suppose X_1, \dots, X_n are independent then

$$\begin{aligned} F(x_1, \dots, x_n) &= \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_n \leq x_n) \\ &= F_1(x_1) \cdots F_n(x_n). \end{aligned}$$

Hence, in the case of independence, $C(u_1, \dots, u_n) = u_1 \cdots u_n$ for all $(u_1, \dots, u_n) \in [0, 1]^n$.

Example 5.2 Suppose (X_1, \dots, X_n) is multivariate standard normal with linear correlation matrix $\boldsymbol{\rho}$. Let $\Phi(z) = \mathbb{P}(Z \leq z)$ for $Z \sim \mathcal{N}(0, 1)$. Then

$$\begin{aligned} F(x_1, \dots, x_n) &= \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \mathbb{P}(F_1(X_1) \leq F_1(x_1), \dots, F_n(X_n) \leq F_n(x_n)) \\ &= C_{\boldsymbol{\rho}}^{Ga}(\Phi(x_1), \dots, \Phi(x_n)), \end{aligned}$$

where

$$C_{\boldsymbol{\rho}}^{Ga}(u_1, \dots, u_n) = \frac{1}{\sqrt{|\boldsymbol{\rho}|(2\pi)^n}} \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_n)} e^{-\frac{1}{2}\mathbf{s}^T \boldsymbol{\rho}^{-1} \mathbf{s}} d\mathbf{s} \quad (28)$$

is called the multivariate Gaussian copula.

Example 5.3 Suppose (X_1, \dots, X_n) is multivariate t with ν degrees of freedom and linear correlation matrix $\boldsymbol{\rho}$ ³⁴. Let $t_\nu(x) = \mathbb{P}(T \leq x)$ where $T \sim t_\nu$. Then

$$\begin{aligned} F(x_1, \dots, x_n) &= \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= \mathbb{P}(F_1(X_1) \leq F_1(x_1), \dots, F_n(X_n) \leq F_n(x_n)) \\ &= C_{\boldsymbol{\rho}}^{t_\nu}(t_\nu(x_1), \dots, t_\nu(x_n)) \end{aligned}$$

where

$$C_{\boldsymbol{\rho}}^{t_\nu}(u_1, \dots, u_n) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})\sqrt{|\boldsymbol{\rho}|(\nu\pi)^n}} \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_n)} \left(1 + \frac{\mathbf{s}^T \boldsymbol{\rho}^{-1} \mathbf{s}}{\nu}\right)^{-\frac{\nu+n}{2}} d\mathbf{s} \quad (29)$$

is called the multivariate t_ν copula.

In Examples 5.2 and 5.3, $|\boldsymbol{\rho}|$ denotes the determinant of the matrix $\boldsymbol{\rho}$. In these examples, the copulas were introduced through the joint distribution, but it is important to remember that the copula characterizes the dependence structure of the multivariate distribution through (26). The Gaussian and t_ν copulas (28) and (29) exist separately from their associated multivariate distributions.

³⁴Its cdf is given by (29) where the upper limits $t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_n)$ are replaced by x_1, \dots, x_n respectively. A multivariate t_ν is easy to generate. Generate a multivariate normal with covariance matrix $\boldsymbol{\Sigma}$ and divide it by $\sqrt{\chi_\nu^2/\nu}$ where χ_ν^2 is an independent chi-squared random variable with ν degrees of freedom.

Example 5.4 The bivariate Gumbel copula C_β^{Gu} is given by

$$C_\beta^{Gu}(u_1, u_2) = \exp \left\{ - \left[(-\ln u_1)^{\frac{1}{\beta}} + (-\ln u_2)^{\frac{1}{\beta}} \right]^\beta \right\}, \quad (30)$$

where $0 < \beta \leq 1$ is a parameter controlling the dependence, $\beta \rightarrow 0^+$ implies perfect dependence (see Section 5.2.3), and $\beta = 1$ implies independence.

Example 5.5 The bivariate Clayton copula C_β^{Cl} is given by

$$C_\beta^{Cl}(u_1, u_2) = (u_1^{-\beta} + u_2^{-\beta} - 1)^{-\frac{1}{\beta}}, \quad (31)$$

where $0 < \beta < \infty$ is a parameter controlling the dependence, $\beta \rightarrow 0^+$ implies independence, and $\beta \rightarrow \infty$ implies perfect dependence. This copula family is sometimes referred to as the Kimeldorf and Sampson family.

Both the Gumbel and Clayton copulas are strict *Archimedean* copulas. Archimedean copulas are defined as follows. Let $\phi : [0, 1] \rightarrow [0, \infty)$ with $\phi(0) = \infty$ and $\phi(1) = 0$ be a continuous, convex, strictly decreasing function. The transformation $\phi^{-1}\phi$ maintains the uniform 1-dimensional distribution since $\phi^{-1}\phi(u) = u$, $u \in [0, 1]$. To obtain a 2-dimensional distribution function use instead of $\phi^{-1}\phi(u)$, $u \in [0, 1]$ the function $\phi^{-1}(\phi(u) + \phi(v))$, $u, v \in [0, 1]$.

Definition 5.2 A strict Archimedean copula with generator ϕ is of the form

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), \quad u, v \in [0, 1]. \quad (32)$$

Example 5.6 The function $\phi(t) = (-\ln t)^{1/\beta}$, $0 < \beta \leq 1$ generates the bivariate Gumbel copula C_β^{Gu} (see Example 5.4).

Example 5.7 The function $\phi(t) = (t^{-\beta} - 1)/\beta$, $\beta > 0$ generates the bivariate Clayton copula C_β^{Cl} (see Example 5.5).

Example 5.8 The function $\phi(t) = -\ln((e^{-\beta t} - 1)/(e^{-\beta} - 1))$, $\beta \in \mathbb{R} \setminus \{0\}$ generates the bivariate Frank copula

$$C_\beta^{Fr}(u, v) = -\frac{1}{\beta} \ln \left(1 + \frac{(e^{-\beta u} - 1)(e^{-\beta v} - 1)}{e^{-\beta} - 1} \right)$$

(see Frank [Fra79]).

If $\phi(0) < \infty$, then the term *strict* in Definition 5.2 is dropped and $\phi^{-1}(s)$ in (32) is replaced by the pseudo-inverse $\phi^{[-1]}(s)$ which equals $\phi^{-1}(s)$ if $0 \leq s \leq \phi(0)$ and is zero otherwise.

Example 5.9 The function $\phi(t) = 1 - t$, $t \in [0, 1]$ satisfies $\phi(0) = 1$ and hence $\phi^{[-1]}(t) = \max(1 - t, 0)$. It generates the non-strict Archimedean copula

$$C(u, v) = \max(u + v - 1, 0).$$

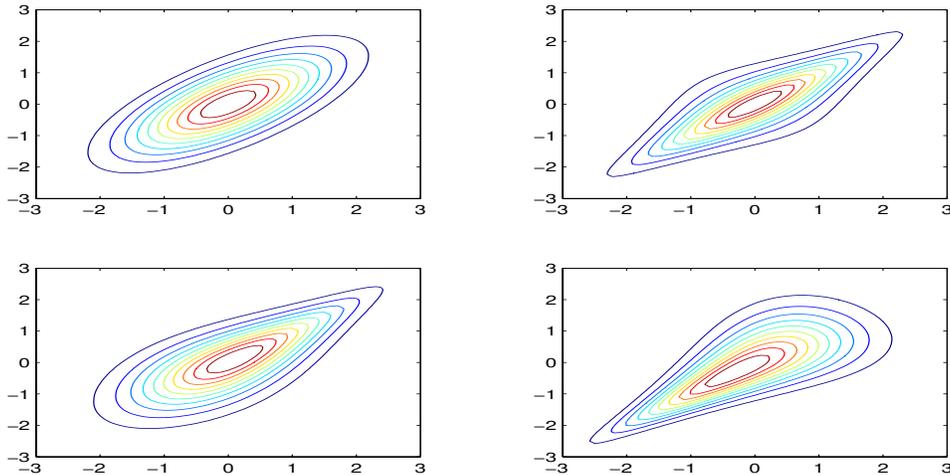


Figure 7: Contours of constant density for different bivariate distributions with standard normal marginals. All have roughly the same linear correlation, and differ only in their copula. Clockwise from upper left: Gaussian, t_2 , Gumbel, Clayton. See Examples 5.2, 5.3, 5.4 and 5.5 for the copula definitions.

The class of Archimedean copulas has many nice properties, including various simple multivariate extensions. For more on Archimedean copulas see [Lin00b], [Nel99], [Joe97] and Embrechts, Lindskog and McNeil [ELM01].

Figure 7 illustrates how the choice of a copula can affect the joint distribution. Each figure shows contours of constant density of a bivariate distribution (X, Y) with standard normal marginals and linear correlations $\rho \approx 0.7$. The differences in the distributions is due to the choice of the copula. (For an introduction on the choice of a copula, see Frees and Valdez [FV98].)

The following theorem provides a bound for the joint cdf.

Theorem 5.2 (Fréchet) *Let F be the joint cdf of distribution with univariate marginals F_1, \dots, F_n . Then for all $\mathbf{x} \in \mathbb{R}^n$.*

$$\underbrace{\max\{0, F_1(x_1) + \dots + F_n(x_n) - (n-1)\}}_{C_L(F_1(x_1), \dots, F_n(x_n))} \leq \underbrace{F(x_1, \dots, x_n)}_{C(F_1(x_1), \dots, F_n(x_n))} \leq \underbrace{\min\{F_1(x_1), \dots, F_n(x_n)\}}_{C_U(F_1(x_1), \dots, F_n(x_n))}.$$

The function $C_U(u_1, \dots, u_n)$ is a copula for all $n \geq 2$, but the function $C_L(u_1, \dots, u_n)$ is a copula for $n = 2$ only. If $n = 2$, the copulas C_L and C_U are the bivariate cdf's of the random vectors $(U, 1 - U)$ and (U, U) respectively, where $U \sim U(0, 1)$.

Another important property of copulas is their invariance under an increasing transformation of the marginals.

Theorem 5.3 *Let X_1, \dots, X_n be continuous random variables with copula C . Let $\alpha_1, \dots, \alpha_n$ be strictly increasing transformations. Then the random vector $(\alpha_1(X_1), \dots, \alpha_n(X_n))$ has the same copula C as (X_1, \dots, X_n) .*

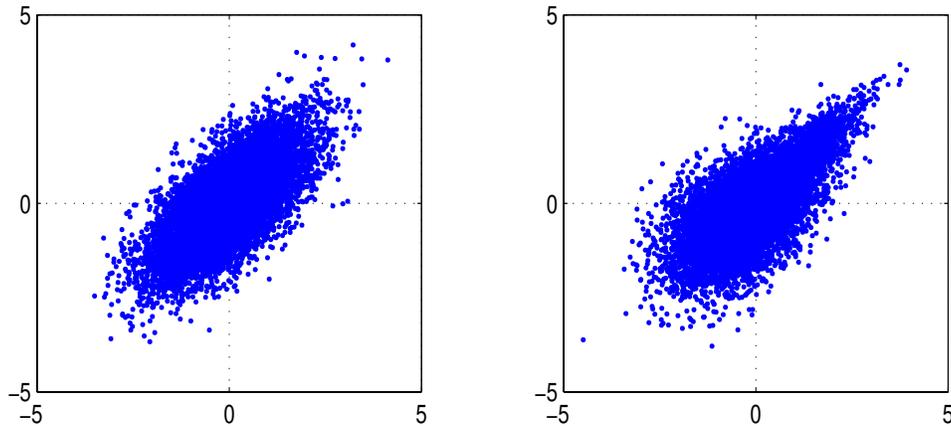


Figure 8: Simulation of 10000 realizations from bivariate distributions both with standard normal marginals and linear correlation of $\rho \approx 0.7$. The distribution on the left has a Gaussian copula, on the right a Gumbel copula. Compare the shapes with those illustrated in Figure 7, where the population distribution is used.

5.2 Measures of Dependence

As already mentioned, linear correlation is the only measure of dependence involved in the mean-variance portfolio theory. This theory assumes, either implicitly or explicitly, that returns are multivariate normal. This assumption seems implausible today given the many complex financial products in the marketplace and the empirical evidence against normality. Without the restrictive assumption of normality, is linear correlation still an appropriate measure of dependence?

Linear correlation is often used in the financial community to describe any form of dependence. As illustrated in [EMS01] and Embrechts, McNeil and Straumann [EMS99], linear correlation is often a very misunderstood measure of dependence. Consider the following example.

Example 5.10 Figure 8 represent 10000 simulations from bivariate distributions $(X, Y)_L$ and $(X, Y)_R$. In both cases X and Y have a standard normal distribution with (approximately) the same linear correlation $\rho \approx 0.7$. Thus, on the basis of the marginal distributions and linear correlation, the two distributions are indistinguishable. The two distributions are however clearly different. If positive values represent losses, the distribution on the right is clearly of greater concern to the risk manager since large losses in X and Y occur simultaneously. The two distributions differ only in their copula.

In the figure on the left the dependence structure is given by the bivariate Gaussian copula. Since the marginals are standard normal, this means that distribution is the bivariate standard normal distribution with the given correlation coefficient. The copula in the figure on the right the Gumbel copula given in (30) with $\beta = 1/2$. Various values of β were tried until the simulation sample linear correlation was $\rho \approx 0.7$.

We now briefly describe several measures of dependence which may be useful to the risk manager. Again the reader is encouraged to look at the above references, especially [EMS01] for details.

5.2.1 Linear Correlation

The linear correlation coefficient ρ , defined in (24), is a commonly misused measure of dependence. To illustrate the confusion involved in interpreting it, consider the following classic example. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and let $Y = X^2$. Then $\rho(X, Y) = 0$, yet clearly X and Y are dependent. Unless we are willing to make certain assumptions about the multivariate distribution, linear correlation can therefore be a misleading measure of dependence. Since the copula of a multivariate distribution describes its dependence structure we would like to use measures of dependence which are copula-based. Linear correlation is not such a measure.

5.2.2 Rank Correlation

Two well-known rank correlation measures which are copula based and have better properties than linear correlation are the Kendall's tau and Spearman's rho.

Definition 5.3 *Let (X_1, Y_1) and (X_2, Y_2) be two independent copies of (X, Y) . Then Kendall's tau, denoted ρ_τ , is given by*

$$\rho_\tau(X, Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

If the marginal distributions F_X and F_Y of X and Y are continuous and if F is the bivariate distribution function of (X, Y) with copula C , then ρ_τ can be expressed in terms of C as follows (see [EMS01]):

$$\rho_\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

Definition 5.4 *Let $X \sim F_X$ and $Y \sim F_Y$. Spearman's correlation, denoted ρ_S , is the linear correlation of $F_X(X)$ and $F_Y(Y)$, that is,*

$$\rho_S(X, Y) = \rho(F_X(X), F_Y(Y)).$$

Spearman's correlation can also be expressed in a form similar to Definition 5.3 (see [Lin00b]). Let (X_1, Y_1) , (X_2, Y_2) and (X_3, Y_3) be three independent copies of (X, Y) . Then

$$\rho_S(X, Y) = 3(\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

If the marginal distributions are continuous, ρ_S is related to the copula of the joint distribution as follows:

$$\rho_S(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3.$$

Whereas linear correlation is a measure of linear dependence, both Kendall's tau and Spearman's rho are measures of monotonic dependence. Since they are copula based, they are invariant under strictly increasing transformations³⁵. Indeed, if α_1, α_2 are strictly increasing transformations, then

$$\begin{aligned} \rho_\tau(\alpha_1(X_1), \alpha_2(X_2)) &= \rho_\tau(X_1, X_2), \\ \rho_S(\alpha_1(X_1), \alpha_2(X_2)) &= \rho_S(X_1, X_2), \\ \text{but } \rho(\alpha_1(X_1), \alpha_2(X_2)) &\neq \rho(X_1, X_2). \end{aligned}$$

³⁵Recall that invariance under increasing transformations is a property of copulas.

5.2.3 Comonotonicity

An additional important property of these rank correlations is their handling of *perfect dependence*. By perfect dependence we mean intuitively that X and Y are monotone functions of the same source of randomness. Recall that in the bivariate case, the Fréchet bounds C_L and C_U in Theorem 5.2 are themselves copulas. The following theorem shows that if the copula is C_L or C_U then X and Y are perfectly dependent.

Theorem 5.4 ([EMS01]). *Suppose that the copula C of (X, Y) is either C_L or C_U . Then there exist monotone functions α and β and a random variable Z such that*

$$(X, Y) \stackrel{d}{=} (\alpha(Z), \beta(Z)).$$

If $C = C_L$ then α and β are increasing and decreasing respectively. If $C = C_U$, then both α and β are increasing.

X and Y are said to be *countermonotonic* if they have copula C_L . If they have copula C_U , they are said to be *comonotonic*. In fact, when F_X and F_Y are continuous,

$$\begin{aligned} C = C_L &\iff Y = T(X) \text{ a.s., } T = F_Y^{-1} \circ (1 - F_X) \searrow, \\ C = C_U &\iff Y = T(X) \text{ a.s., } T = F_Y^{-1} \circ F_X \nearrow. \end{aligned}$$

Kendall's tau and Spearman's rho handle perfect dependence in a reasonable manner. Indeed,

Theorem 5.5 ([EMS01]). *Let $(X, Y) \sim F$ with continuous marginals and copula C . Then*

$$\begin{aligned} \rho_\tau(X, Y) = -1 &\iff \rho_S(X, Y) = -1 \iff C = C_L \iff X \text{ and } Y \text{ are countermonotonic,} \\ \rho_\tau(X, Y) = 1 &\iff \rho_S(X, Y) = 1 \iff C = C_U \iff X \text{ and } Y \text{ are comonotonic.} \end{aligned}$$

The following theorem due to Höfding and Fréchet deals with linear correlation. See [EMS01] for its proof.

Theorem 5.6 *Let (X, Y) be a random vector with marginals non-degenerate F_X and F_Y and unspecified dependence structure. If X and Y have finite variance, then*

1. *The set of possible linear correlations is a closed interval $[\rho_{min}, \rho_{max}]$ with $\rho_{min} < 0 < \rho_{max}$.*
2. *The extremal linear correlation $\rho = \rho_{min}$ is attained iff X and Y are countermonotonic; $\rho = \rho_{max}$ is attained iff X and Y are comonotonic.*
3. *$\rho_{min} = -1 \iff X$ and $-Y$ are of the same type³⁶; $\rho_{max} = 1 \iff X$ and Y are of the same type.*

The following example shows that linear correlation does not handle perfect dependence in a reasonable manner.

³⁶Recall that two random variables are the same type if their distributions are the same up to a change in location and scale.

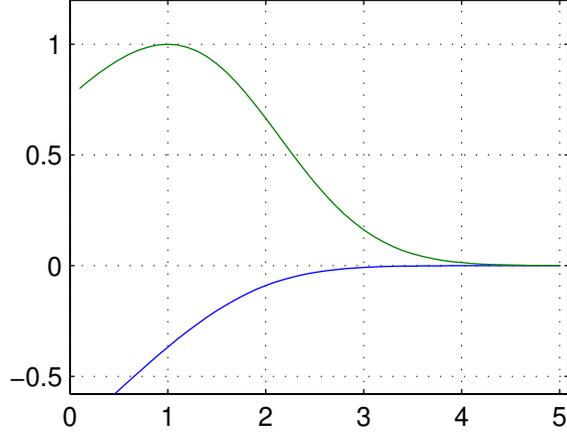


Figure 9: Range of maximal and minimal linear correlation in Example 5.11. The x -axis is in units of σ . As σ increases, both the maximal and minimal linear correlations tend to zero.

Example 5.11 ([EMS01]). Let $X \sim \text{Lognormal}(0, 1)$ and $Y \sim \text{Lognormal}(0, \sigma^2)$ with $\sigma > 0$. By Theorem 5.6, $\rho = \rho_{min}$ and $\rho = \rho_{max}$ when X and Y are countermonotonic and comonotonic respectively. By Theorem 5.4, $(X, Y) \stackrel{d}{=} (\alpha(Z), \beta(Z))$, and in fact, $(X, Y) \stackrel{d}{=} (e^Z, e^{-\sigma Z})$ when X and Y are countermonotonic and $(X, Y) \stackrel{d}{=} (e^Z, e^{\sigma Z})$ when X and Y are comonotonic, where $Z \sim \mathcal{N}(0, 1)$. Hence $\rho_{min} = \rho(e^Z, e^{-\sigma Z})$ and $\rho_{max} = \rho(e^Z, e^{\sigma Z})$ where $Z \sim \mathcal{N}(0, 1)$. Using the properties of the lognormal distribution, these maximal and minimal correlations can be evaluated explicitly and one gets

$$\rho_{min} = \frac{e^{-\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}} \quad , \quad \rho_{max} = \frac{e^{\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}.$$

As σ increases, the maximal and minimal linear correlation both tend to zero even though X and Y are monotonic functions of the same source of randomness. This is illustrated in Figure 9.

5.2.4 Tail Dependence

There is a saying in finance that in times of stress all correlations go to one³⁷. While it shows that the financial community uses linear correlation to describe any measure of dependence, it can also serve as motivation for the next measure of dependence, known as tail dependence.

Bivariate tail dependence measures the amount of dependence in the upper and lower quadrant tail of the distribution. This is of great interest to the risk manager trying to guard against concurrent bad events in the tails.

³⁷See [CPB00] for example.

Definition 5.5 Let $X \sim F_X$ and $Y \sim F_Y$ and observe that as $\alpha \rightarrow 1^-$, $F_X^{-1}(\alpha) \rightarrow \infty$ and $F_Y^{-1}(\alpha) \rightarrow \infty$. The coefficient of upper tail dependence λ_U is

$$\lambda_U(X, Y) = \lim_{\alpha \rightarrow 1^-} \mathbb{P}(Y > F_Y^{-1}(\alpha) | X > F_X^{-1}(\alpha)) \quad (33)$$

provided the limit exists. If $\lambda_U = 0$, then X and Y are said to asymptotically independent in the upper tail. If $\lambda_U \in (0, 1]$, then X and Y are asymptotically dependent in the upper tail. The coefficient of lower tail dependence λ_L is similarly defined:

$$\lambda_L(X, Y) = \lim_{\alpha \rightarrow 0^+} \mathbb{P}(Y < F_Y^{-1}(\alpha) | X < F_X^{-1}(\alpha)).$$

Since

$$\lambda_U(X, Y) = \lim_{\alpha \rightarrow 1^-} \frac{1 - \mathbb{P}(X \leq F_X^{-1}(\alpha)) - \mathbb{P}(Y \leq F_Y^{-1}(\alpha)) + \mathbb{P}(X \leq F_X^{-1}(\alpha), Y \leq F_Y^{-1}(\alpha))}{1 - \mathbb{P}(X \leq F_X^{-1}(\alpha))},$$

λ_U , as well as λ_L , can be expressed in terms of copulas. Let (X, Y) have continuous distribution F with copula C . It is easily seen that the coefficient of upper tail dependence λ_U can be expressed as

$$\lambda_U(X, Y) = \lim_{\alpha \rightarrow 1^-} \frac{\bar{C}(\alpha, \alpha)}{1 - \alpha}, \quad (34)$$

where $\bar{C}(\alpha, \alpha) = 1 - 2\alpha + C(\alpha, \alpha)$ ³⁸. Similarly,

$$\lambda_L(X, Y) = \lim_{\alpha \rightarrow 0^+} \frac{C(\alpha, \alpha)}{\alpha}.$$

Example 5.12 Recall the simulation Example 5.10. In this example, both distributions had the same marginal distributions with the same linear correlation. Yet the distributions were clearly different in the upper tail. This difference came from the choice of copula and may now be quantified by using the notion of upper tail dependence. In Figure 8 on the left, $F(x, y) = C_\rho^{Ga}(\Phi(x), \Phi(y))$, Φ denotes the standard $\mathcal{N}(0, 1)$ cdf and C_ρ^{Ga} is given by (28) that is, the distribution is a bivariate standard normal with linear correlation $\rho = 0.7$. The coefficient of upper tail dependence can be calculated explicitly³⁹,

$$\lambda_U(X, Y) = 2 \lim_{x \rightarrow \infty} \bar{\Phi} \left(\frac{x\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) = 0,$$

which is a general characteristic of Gaussian copulas. This means that if we go far enough out into the tail then extreme events occur independently in X and Y . In the figure of the right, $F(x, y) = C_\beta^{Gu}(\Phi(x), \Phi(y))$, with C_β^{Gu} given by (30), where the dependence parameter β was chosen to give (approximately) the same linear correlation⁴⁰. In the case

³⁸If $(U_1, U_2)^T \sim C$ then

$$\bar{C}(u_1, u_2) = \mathbb{P}(U_1 > u_1, U_2 > u_2) = 1 - u_1 - u_2 + C(u_1, u_2).$$

³⁹ $\bar{\Phi}(x) = 1 - \Phi(x)$, and, below $\bar{t}_\nu(x) = 1 - t_\nu(x)$.

⁴⁰The dependence parameter β of the bivariate Gumbel copula is related to Kendall's tau by $\rho_\tau = 1 - \beta$.

of the Gumbel copula a simple calculation shows that for all $0 < \beta < 1$, the coefficient of upper tail dependence is

$$\lambda_U(X, Y) = 2 - 2^\beta.$$

Hence, for the Gumbel copula, $\lambda_U \neq 0$ for $0 < \beta < 1$.

Suppose the risk manager tries to account for heavy tails of a distribution by simply modelling the joint distribution as a multivariate t_ν . He will not get $\lambda_U = 0$ as in the case of the multivariate normal distribution.

Example 5.13 If $(X, Y) \sim t_\nu$ with any linear correlation $\rho \in (-1, 1)$ then it can be shown ([EMS01]) that

$$\lambda_U(X, Y) = 2\bar{t}_{\nu+1} \left(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right).$$

Hence for all $\rho \in (-1, 1)$ there is upper tail dependence of the bivariate t_ν . The stronger the linear correlation and the lower the degrees of freedom, the stronger the upper tail dependence.

5.3 Elliptical Distributions

There are distributions other than multivariate normal where linear correlation can be used effectively. These are the spherical, or more generally, the elliptical distributions. Elliptical distributions extend in a natural way the class of multivariate normal distributions. Linear correlation (when it exists) will still be the canonical measure of dependence, yet elliptical distributions can display heavy tails.

We shall define first the spherical distributions. These extend the class of standard multivariate normal distributions with zero correlations ([FKN90], [EMS01]).

Definition 5.6 The random vector $\mathbf{X} \in \mathbb{R}^n$ is said to be spherically distributed if

$$\Gamma \mathbf{X} \stackrel{d}{=} \mathbf{X} \quad \forall \quad \Gamma \in \mathcal{O}(n)$$

where $\mathcal{O}(n)$ is the group of $n \times n$ orthogonal matrices.

In other words, the distribution of \mathbf{X} is invariant under rotation of the coordinates. Here are further characterizations.

Theorem 5.7 The random vector $\mathbf{X} \in \mathbb{R}^n$ has a spherical distribution iff its characteristic function $\Psi_{\mathbf{X}}$ satisfies one of the following equivalent conditions:

1. $\Psi_{\mathbf{X}}(\Gamma^T \mathbf{t}) = \Psi_{\mathbf{X}}(\mathbf{t}) \quad \forall \quad \Gamma \in \mathcal{O}(n)$;
2. There exists a function $\phi(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that $\Psi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^T \mathbf{t})$, that is, $\Psi_{\mathbf{X}}(\mathbf{t}) = \phi(\sum_{i=1}^n t_i^2)$,

where $\mathbf{t} = (t_1, \dots, t_n)$. Alternatively, spherical distributions admit a stochastic representation, namely, $\mathbf{X} \in \mathbb{R}^n$ has a spherical distribution iff there exists a non-negative random variable R and random vector \mathbf{U} independent of R and uniformly distributed over the unit hypersphere $S_n = \{\mathbf{s} \in \mathbb{R}^n \mid \|\mathbf{s}\| = 1\}$ such that

$$\mathbf{X} \stackrel{d}{=} RU. \tag{35}$$

Example 5.14 Let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ then

$$\Psi_{\mathbf{X}}(\mathbf{t}) = e^{-\frac{1}{2}(\mathbf{t}^T \mathbf{t})} = e^{-\frac{1}{2}(\sum_{i=1}^n t_i^2)},$$

and so $\phi(u) = e^{-u/2}$. Additionally, $R \sim \sqrt{\chi_n^2}$ in the stochastic representation (35).

The function ϕ is called the *characteristic generator* of the spherical distribution. We write

$$\mathbf{X} \sim S_n(\phi)$$

to indicate that $\mathbf{X} \in \mathbb{R}^n$ is spherically distributed with generator ϕ . Note that if \mathbf{X} possesses a density, then Theorem 5.7 requires that it is of the form

$$f(\mathbf{x}) = g(\mathbf{x}^T \mathbf{x}) = g\left(\sum_{i=1}^n x_i^2\right)$$

for some non-negative function g . The curves of constant density are spheroids in \mathbb{R}^n .

Example 5.15 If $\mathbf{X} \in \mathbb{R}^n$ has a multivariate t_ν distribution with zero correlation, then

$$f(\mathbf{x}) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}} \left(1 + \frac{\mathbf{x}^T \mathbf{x}}{\nu}\right)^{-\frac{\nu+n}{2}}.$$

\mathbf{X} is therefore spherically distributed.

Table 2 gives a partial list of the spherical distributions used in finance.

Type	pdf $f(\mathbf{x})$ or ch.f. $\Psi(\mathbf{t})$
Normal	$f(\mathbf{x}) = c \exp(-\mathbf{x}^T \mathbf{x}/2)$
t_ν	$f(\mathbf{x}) = c(1 + \mathbf{x}^T \mathbf{x}/\nu)^{-(\nu+n)/2}$
Logistic	$f(\mathbf{x}) = c \exp(-\mathbf{x}^T \mathbf{x})/[1 + \exp(-\mathbf{x}^T \mathbf{x})]^2$
Scale Mixture	$f(\mathbf{x}) = c \int_0^\infty t^{-n/2} \exp(-\mathbf{x}^T \mathbf{x}/2t) dG(t)$, $G(t)$ a c.d.f.
Stable Laws	$\Psi(\mathbf{t}) = \exp\{r(\mathbf{t}^T \mathbf{t})^{\alpha/2}\}$, $0 < \alpha \leq 2$ and $r > 0$

Table 2: Partial list of spherical distributions used in finance.

Recall that if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, then $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{X}$ has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$. Elliptical distributions are defined from spherical distributions in a similar manner. They are affine transformations of spherical distributions.

Definition 5.7 Let $\mathbf{X} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$. Then \mathbf{X} has an elliptical distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ if

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$$

where $\mathbf{Y} \sim S_k(\phi)$, and $\mathbf{A} \in \mathbb{R}^{n \times k}$, $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$, with $\text{rank}(\boldsymbol{\Sigma}) = k$.

Since the characteristic function of \mathbf{X} may be written

$$\Psi_{\mathbf{X}}(\mathbf{t}) = e^{it^T \boldsymbol{\mu}} \phi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}),$$

we use the notation

$$\mathbf{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi).$$

In this representation only $\boldsymbol{\mu}$ is uniquely determined. Since both $\boldsymbol{\Sigma}$ and ϕ are determined up to a positive constant $\boldsymbol{\Sigma}$ may be chosen to be the covariance matrix if variances are finite (which we assume here). An elliptically distributed random variable $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ is thus described by its mean, covariance matrix and its characteristic generator. If \mathbf{X} possesses a density, then it is of the form

$$f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})) \quad (36)$$

so that contours of constant density are ellipsoids in \mathbb{R}^n ⁴¹.

The following theorem describes some properties of linear combinations, marginal distributions and conditional distributions of elliptical distributions.

Theorem 5.8 ([FKN90]) *Let $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.*

1. *If $\mathbf{B} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{\nu} \in \mathbb{R}^m$, then*

$$\boldsymbol{\nu} + \mathbf{B}\mathbf{X} \sim E_m(\boldsymbol{\nu} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T, \phi).$$

Hence any linear combination of elliptically distributed variates is elliptical with the same characteristic generator.

2. *Partition \mathbf{X} , $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ into*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\mathbf{X}^{(1)} \in \mathbb{R}^m$, $\boldsymbol{\mu}^{(1)} \in \mathbb{R}^m$ and $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{m \times m}$, $0 < m < n$. Then

$$\mathbf{X}^{(1)} \sim E_m(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}, \phi) \quad , \quad \mathbf{X}^{(2)} \sim E_{n-m}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}, \phi).$$

Hence all marginals of an elliptical distribution are also elliptical with the same generator.

3. *Partition \mathbf{X} , $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ as above and assume that $\boldsymbol{\Sigma}$ is strictly positive definite. Then*

$$\mathbf{X}^{(1)} | \mathbf{X}^{(2)} = \mathbf{x}_0^{(2)} \sim E_m(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}, \tilde{\phi}),$$

where

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_0^{(2)} - \boldsymbol{\mu}^{(2)}) \quad , \quad \boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}.$$

Hence the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)}$ is also elliptical, though with different generator⁴².

⁴¹For example if $\text{rank}(\boldsymbol{\Sigma}) = n$ and \mathbf{Y} has density of the form $g(\mathbf{y}^T \mathbf{y})$.

⁴²The form of the generator $\tilde{\phi}$ can be related to ϕ through the stochastic representation of an elliptically distributed random vector in (35). See [FKN90] for details.

The importance of the class of elliptical distributions to risk management can be seen in the following theorem. It indicates that the standard approaches to risk management apply to a linear portfolio with elliptically distributed risk factors.

Theorem 5.9 ([EMS01]) *Suppose $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ with finite variances for all univariate marginals. Let*

$$\mathcal{P} = \left\{ Z = \sum_{i=1}^n w_i X_i \mid w_i \in \mathbb{R} \right\}$$

be the set of all linear portfolios. Then:

1. (Subadditivity of VaR.) For any two portfolios $Z_1, Z_2 \in \mathcal{P}$ and $0.5 \leq \alpha < 1$,

$$\text{VaR}_\alpha(Z_1 + Z_2) \leq \text{VaR}_\alpha(Z_1) + \text{VaR}_\alpha(Z_2).$$

2. (Equivalence of variance and any other positive homogeneous risk measure.) Let ρ be any real valued, positive homogeneous risk measure depending only on the distribution of a random variable X . Then for $Z_1, Z_2 \in \mathcal{P}$,

$$\rho(Z_1 - \mathbb{E}Z_1) \leq \rho(Z_2 - \mathbb{E}Z_2) \iff \sigma_{Z_1}^2 \leq \sigma_{Z_2}^2.$$

3. (Markowitz risk minimizing portfolio.) Let ρ be as in 2., but also translation invariant, and let

$$\mathcal{E} = \left\{ Z = \sum_{i=1}^n w_i X_i \mid w_i \in \mathbb{R}, \sum_{i=1}^n w_i = 1, \mathbb{E}Z = r \right\}$$

be the subset of portfolios with the same expected return r . Then

$$\arg \min_{Z \in \mathcal{E}} \rho(Z) = \arg \min_{Z \in \mathcal{E}} \sigma_Z^2.$$

The theorem⁴³ states that:

⁴³Because of the importance of Theorem 5.9 and because its proof is illuminating and straightforward we shall sketch it. It is based on the observation that (Z_1, Z_2) is elliptical and so portfolios Z_1, Z_2 and $Z_1 + Z_2$ are all of the same type. Let q_α , $1/2 < \alpha < 1$, denote the α quantile of the corresponding standardized distribution. Then

$$\begin{aligned} \text{VaR}_\alpha(Z_1) &= \mathbb{E}Z_1 + \sigma_{Z_1} q_\alpha \\ \text{VaR}_\alpha(Z_2) &= \mathbb{E}Z_2 + \sigma_{Z_2} q_\alpha \\ \text{VaR}_\alpha(Z_1 + Z_2) &= \mathbb{E}Z_1 + \mathbb{E}Z_2 + \sigma_{Z_1+Z_2} q_\alpha \end{aligned}$$

but $\sigma_{Z_1+Z_2} \leq \sigma_{Z_1} + \sigma_{Z_2}$ and $q_\alpha > 0$, proving (1.). Next, note that there exists $a > 0$ such that $Z_1 - \mathbb{E}Z_1 \stackrel{d}{=} a(Z_2 - \mathbb{E}Z_2)$, so that $a \leq 1 \iff \sigma_{Z_1}^2 \leq \sigma_{Z_2}^2$. Since the risk measure ρ is assumed positive homogeneous and depends only on the distribution of Z ,

$$\rho(Z_1 - \mathbb{E}Z_1) = \rho(a(Z_2 - \mathbb{E}Z_2)) = a\rho(Z_2 - \mathbb{E}Z_2)$$

and hence

$$\rho(Z_1 - \mathbb{E}Z_1) \leq \rho(Z_2 - \mathbb{E}Z_2) \iff a \leq 1 \iff \sigma_{Z_1}^2 \leq \sigma_{Z_2}^2 \tag{37}$$

which proves (2.). Now consider only portfolios in \mathcal{E} . Then (37) holds with $\mathbb{E}Z_1 = \mathbb{E}Z_2 = r$. However, using translation invariance of ρ , $\rho(Z_j - r) = \rho(Z_j) - r$ for $j = 1, 2$. This gives

$$\rho(Z_1) \leq \rho(Z_2) \iff \sigma_{Z_1}^2 \leq \sigma_{Z_2}^2$$

proving (3.).

- For any linear portfolio of elliptical risk factors, VaR is a coherent measure of risk.
- If the risk factors are elliptical, the linear correlation is the canonical measure of dependence.
- For elliptical risk factors, the Markowitz mean variance optimal portfolio, for a given level of expected return, will be the same regardless of whether the risk measure is given by the variance, VaR , expected shortfall or any other positive homogeneous, translation invariant risk measure. Hence, all the usual techniques of portfolio theory and risk management apply.
- It may be strange at first that the expected shortfall $S_\alpha(X)$, for example, which does not involve subtraction of the mean (see (22)), can be used instead of the variance in Markowitz' risk minimization portfolio theory. This is because one considers a set of portfolios \mathcal{E} , all of the same mean. Since $S_\alpha(X - \mathbb{E}X) = S_\alpha(X) - \mathbb{E}X$ and since $\mathbb{E}X$ is the same for all portfolios X in \mathcal{E} , the term $\mathbb{E}X$ can be ignored.

Note that elliptical distributions are not required to be thin-tailed. The multivariate normal is but one elliptical distribution. The risk manager may well feel that the risk factors under consideration are better modelled using a heavy-tailed elliptical distribution⁴⁴. The usual techniques then apply, but the risk of a linear portfolio will be greater than if the risk factors were assumed multivariate normal.

6 Univariate Extreme Value Theory

Managing extreme market risk is a goal of any financial institution or individual investor. In an effort to guarantee solvency, financial regulators require most financial institutions to maintain a minimum level of capital in reserve. The recommendation of the Basle Committee [oBS5b] of a minimum capital reserve requirement based on VaR is an attempt to manage extreme market risks. Recall that VaR is nothing more than a quantile of a probability distribution. The minimum capital reserve is then a multiple of this high quantile, usually computed with $\alpha = 0.99$. Therefore it is very important to attempt to model correctly the tail of probability distribution of returns (profit and losses). The primary difficulty is that we are trying to model events about which we know very little. By definition, these events are rare. The model must allow for these rare but very damaging events. Extreme value theory (EVT) approaches the modelling of these rare and damaging events in a statistically sound way. Once the risks have been modelled they may be measured. We will use VaR and *Expected Shortfall* to measure them.

Extreme value theory (EVT) has its roots in hydrology, where, for example, one needed to compute how high a sea dyke had to be to guard against a 100 year storm. EVT has recently found its way into the financial community. The reader interested in a solid background may now consult various texts on EVT such as Embrechts, Klüppelberg, and Mikosch [EKM97], Reiss and Thomas [RT01] and Beirlant, Teugels and Vynckier [BTV96]. For discussions of the use of EVT in risk management, see Embrechts [Emb00] and Diebold, Schuermann and Stroughair [DSS00].

The modelling of extremes may be done in two different ways: modelling the *maximum* of a collection of random variables, and modelling the *largest values* over some high threshold. We start, for historical reasons, with the first method, called block maxima.

⁴⁴In a recent paper, Lindskog [Lin00a] compares estimators for linear correlation showing that the standard covariance estimator (17) performs poorly for heavy-tailed elliptical data. Several alternatives are proposed and compared.

6.1 Limit Law for Maxima

The Fisher-Tippett theorem is one of two fundamental theorems in EVT. It does for the maxima of *i.i.d.* random variables what the central limit theorem does for sums. It provides the limit law for maxima.

Theorem 6.1 (*Fisher-Tippett(1928)*). *Let (X_n) be a sequence of i.i.d. random variables with distribution F . Let $M_n = \max(X_1, \dots, X_n)$. If there exist norming constants $c_n > 0$ and $d_n \in \mathbb{R}$ and some non-degenerate distribution function H such that*

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H,$$

then H is one of the following three types:

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

The distributions Φ_α, Ψ_α , and Λ are called standard extreme value distributions. The expressions given above are cumulative distribution functions. The Weibull is usually defined as having support $(0, \infty)$ but, in the context of extreme value theory, it has support on $(-\infty, 0)$, as indicated in the theorem. These distributions are related:

$$X \sim \Phi_\alpha \iff \ln X^\alpha \sim \Lambda \iff -1/X \sim \Psi_\alpha.$$

A one-parameter representation of these distributions (due to Jenkinson and von Mises) will be useful. The reparameterized version is called the generalized extreme value (GEV) distribution.

$$H_\xi(x) = \begin{cases} \exp\{-(1 + \xi x)^{-1/\xi}\} & : \xi \neq 0 \\ \exp\{-e^{-x}\} & : \xi = 0 \end{cases}$$

where $1 + \xi x > 0$. The standard extreme value distributions Φ_α, Ψ_α , and Λ follow by taking $\xi = \alpha^{-1} > 0$, $\xi = -\alpha^{-1} < 0$, and $\xi = 0$ respectively⁴⁵. Their densities are sketched in Figure 10. The parameter ξ is the *shape parameter* of H . Since for any random variable $X \sim F_X$ and constants $\mu \in \mathbb{R}$ and $\sigma > 0$, the distribution function of $\tilde{X} = \mu + \sigma X$ is given by $F_{\tilde{X}}(x) = F_X(\frac{x-\mu}{\sigma})$, we can add location and scale parameters to the above parameterization, and consider

$$H_{\xi, \mu, \sigma}(x) = H_\xi\left(\frac{x - \mu}{\sigma}\right).$$

⁴⁵Consider, for example, the Fréchet distribution where $\xi = \alpha^{-1} > 0$. Since the support of H_ξ is $1 + \xi x > 0$, one has

$$H_{\alpha^{-1}}(x) = \exp\{-(1 + \alpha^{-1}x)^{-\alpha}\} = \Phi_\alpha(1 + \alpha^{-1}x)$$

for $1 + \alpha^{-1}x > 0$.

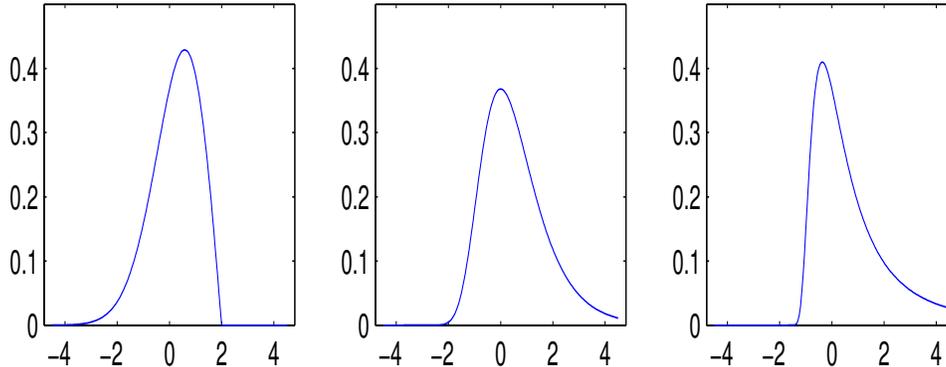


Figure 10: Densities of the generalized extreme value distribution H_ξ . (Left) Weibull with $\xi = -0.5$. (Middle) Gumbel with $\xi = 0$. (Right) Fréchet with $\xi = 0.5$.

If the Fisher-Tippett theorem holds, then we say that F is in the maximum domain of attraction of H and write $F \in MDA(H)$. Most distributions in statistics are in $MDA(H_\xi)$ for some ξ . If $F \in MDA(H_\xi)$ and $\xi = 0$ or $F \in MDA(H_\xi)$ and $\xi < 0$, then F is said to be thin-tailed or short-tailed respectively. Thin-tailed distributions ($\xi = 0$) include the normal, exponential, gamma and lognormal. Short-tailed distributions ($\xi < 0$) have a finite right-hand end point and include the uniform and beta distributions. The heavy-tailed distributions, those in the domain of attraction of the Fréchet distribution, $F \in MDA(H_\xi)$, for $\xi > 0$, are of particular interest in finance. They are characterized in the following theorem due to Gnedenko.

Theorem 6.2 (Gnedenko(1943)) *The distribution function $F \in MDA(H_\xi)$ for $\xi > 0$ if and only if $\bar{F}(x) = 1 - F(x) = x^{-1/\xi}L(x)$ for some slowly varying function L ⁴⁶.*

Distributions such as the Student- t , α -stable and *Pareto* are in this class. Note that if $X \sim F$ with $F \in MDA(H_\xi)$, $\xi > 0$ then all moments $\mathbb{E}X^\beta$ are infinite for $\beta > 1/\xi$. Note also that $\xi < 1$ corresponds to $\alpha > 1$, where α is as in Theorem 6.1.

6.2 Block Maxima Method

We now explain the *block maxima* method, where one assumes in practice that the maximum is distributed as $H_{\xi,\mu,\sigma}$. The implementation of this method requires a great deal of data. Let X_1, X_2, \dots, X_{mn} be daily (negative) returns and divide them into m adjacent blocks of size n . Choose the block size n large enough so that our limiting theorem results apply to $M_n^{(j)} = \max(X_{(j-1)n+1}, \dots, X_{(j-1)n+n})$ for $j = 1, \dots, m$. Our data set must then be long enough to allow for m blocks of length n . There are three parameters, ξ , μ and σ , which need to be estimated, using for example maximum likelihood based on the extreme value distribution. The value of m must be sufficiently large as well, to allow for

⁴⁶The function L is said to be slowly varying (at infinity) if

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, \forall t > 0.$$

a reasonable confidence in the parameter estimation. This is the classic bias-variance trade-off since for a finite data set, increasing the number of blocks m , which reduces the variance, decreases the block size n , which increases the bias. Once the GEV model $H_{\xi,\mu,\sigma}$ is fit using $M_n^{(1)}, \dots, M_n^{(m)}$, we may estimate quantities of interest.

For example, assuming $n = 261$ trading days per year, we may want to find $R_{261,k}$, the daily loss we expect to be exceeded in one year every k years⁴⁷. If this loss is exceeded in a given day, this day is viewed as an exceedance day and the year to which the day belongs is regarded as an exceedance year. While an exceedance year has at least one exceedance day, we are not concerned here with the total number of exceedance days in that year. This would involve taking into consideration the propensity of extremes to form clusters. Since we want M_{261} to be less than $R_{261,k}$ for $k - 1$ of k years, $R_{261,k}$ is the $1 - 1/k$ quantile of M_{261} :

$$R_{261,k} = \inf \left\{ r \mid \mathbb{P}(M_{261} \leq r) \geq 1 - \frac{1}{k} \right\}. \quad (38)$$

If we assume that M_{261} has approximately the $H_{\xi,\mu,\sigma}$ distribution, the quantile $R_{261,k}$ is given by

$$R_{261,k} = H_{\xi,\mu,\sigma}^{-1} \left(1 - \frac{1}{k} \right) \quad (39)$$

$$= \mu + \frac{\sigma}{\xi} \left(\left(-\ln \left(1 - \frac{1}{k} \right) \right)^{-\xi} - 1 \right), \quad \xi \neq 0, \quad (40)$$

since the inverse function of $y = \exp\{-(1 + \xi x)\}^{-1/\xi}$ is $x = (1/\xi)[(-\ln y)^{-\xi} - 1]$. Confidence intervals for $R_{261,k}$ may also be constructed using *profile log-likelihood* functions. The idea is as follows. The GEV distribution $H_{\xi,\mu,\sigma}$ depends on three parameters. Substitute $R_{261,k}$ for μ using (40) and denote the reparameterized H as $H_{\xi,R_{261,k},\sigma}$ after some abuse of notation. Then obtain the log-likelihood $L(\xi, R_{261,k}, \sigma \mid M_1, \dots, M_m)$ for our m observations from $H_{\xi,R_{261,k},\sigma}$. Take $H_0 : R_{261,k} = r$ as the null hypothesis in an asymptotic likelihood ratio test and let $\Theta_0 = (\xi \in \mathbb{R}, R_{261,k} = r, \sigma \in \mathbb{R}^+)$ and $\Theta = (\xi \in \mathbb{R}, R_{261,k} \in \mathbb{R}, \sigma \in \mathbb{R}^+)$ be the constrained and unconstrained parameter spaces respectively. Then under certain regularity conditions we have that

$$-2 \left[\sup_{\Theta_0} L(\theta \mid M_1, \dots, M_m) - \sup_{\Theta} L(\theta \mid M_1, \dots, M_m) \right] \sim \chi_1^2$$

as $m \rightarrow \infty$ where $\theta = (\xi, R_{261,k}, \sigma)$ and χ_1^2 is a chi-squared distribution with one degree of freedom. Let $L(\hat{\xi}, r, \hat{\sigma}) = \sup_{\Theta_0} L(\theta \mid M_1, \dots, M_m)$ and $L(\hat{\xi}, \hat{R}_{261,k}, \hat{\sigma}) = \sup_{\Theta} L(\theta \mid M_1, \dots, M_m)$ denote the constrained and unconstrained maximum log-likelihood values respectively. The α confidence interval for $R_{261,k}$ is the set

$$\{r : L(\hat{\xi}, r, \hat{\sigma}) \geq L(\hat{\xi}, \hat{R}_{261,k}, \hat{\sigma}) - \frac{1}{2}\chi_1^2(\alpha)\},$$

that is, the set r for which the null hypothesis cannot be rejected for level α . See McNeil [McN98a] or K ellezi and Gilli [KG00] for details.

⁴⁷Note the obvious hydrological analogy: How high to build a sea dyke to guard against a k year storm.

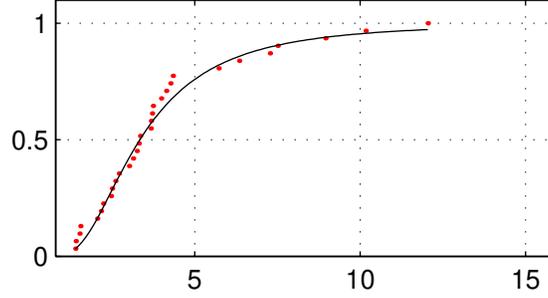


Figure 11: The GEV distribution $H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}$ fitted using the 31 annual maxima of daily (negative, as %) *NASDAQ* returns.

Example 6.1 We have 7570 data points for the *NASDAQ*, which we subdivided into $m = 31$ blocks of roughly $n = 261$ trading days. (The last block, which corresponds to January 2001, has relatively few trading days, but was included because of the large fluctuations.) Estimating the GEV distribution by maximum likelihood leads to $\hat{\xi} = .319$, $\hat{\mu} = 2.80$ and $\hat{\sigma} = 1.38$. The value of $\hat{\xi}$ corresponds to $\hat{\alpha} = 1/\hat{\xi} = 3.14$, which is in the expected range for financial data. The GEV fit is not perfect (see Figure 11). Choosing $k = 20$ yields an estimate of the twenty year return level $\hat{R}_{261,20} = 9.62\%$. Figure 12, which displays the log-likelihood corresponding to the null-hypothesis that $\hat{R}_{261,20} = r$, where r is displayed on the abscissa, also provides the corresponding confidence interval.

6.3 Using the Block Maxima Method for Stress Testing

For the purpose of *stress testing* (worst case scenario), it is the high quantiles of the daily return distribution F that we are interested in, not those of M_n . If the $X_i \sim F$ have a continuous distribution, we have

$$\mathbb{P}(M_n \leq R_{n,k}) = 1 - 1/k.$$

If they are also *i.i.d.*,

$$\mathbb{P}(M_n \leq R_{n,k}) = (\mathbb{P}(X \leq R_{n,k}))^n,$$

where $X \sim F$, and hence

$$\mathbb{P}(X \leq R_{n,k}) = (1 - 1/k)^{1/n}. \quad (41)$$

This means that $R_{n,k}$ is the $(1 - 1/k)^{1/n}$ quantile of the marginal distribution F . Suppose we would like to calculate VaR at very high quantiles for the purposes of stress testing. The block size n has been fixed for the calibration of the model. This leaves the parameter k for the $R_{n,k}$ return level free. High α quantiles, $x_\alpha = F^{-1}(\alpha)$, of F may then be computed from (41) by choosing $\alpha = (1 - 1/k)^{1/n}$, that is $k = 1/(1 - \alpha^n)$. Hence

$$VaR_\alpha(X) = R_{n,k} \quad \text{where} \quad k = \frac{1}{1 - \alpha^n}. \quad (42)$$

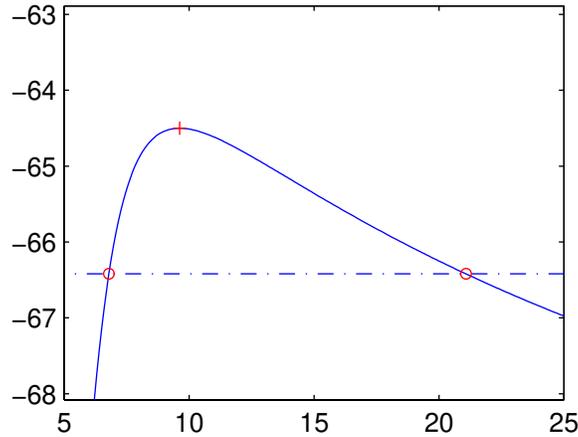


Figure 12: The profile log-likelihood curve for the 20 year return level $R_{261,20}$ for *NASDAQ*. The abscissa displays return levels (as %) and the ordinate displays log-likelihoods. The point estimate $\hat{R}_{261,20} = 9.62\%$ corresponds to the location of the maximum and the asymmetric 95% confidence interval, computed using the profile log-likelihood curve, is (6.79%, 21.1%).

For the *NASDAQ* data, our choice of $k = 20$, corresponds to $\alpha = .9998$, and $\widehat{VaR}_{\alpha=.9998}(X) = \hat{R}_{261,20} = 9.62\%$. In practice α is given, and one chooses $k = 1/(1 - \alpha^n)$, then computes $R_{n,k}$ using (40) and thus one obtains $VaR_{\alpha}(X) = R_{n,k}$.

We assumed independence but, in finance, this assumption is not realistic. At best, the marginal distribution F can be viewed as stationary. For the extension of the Fisher-Tippett theorem to stationary time series see Leadbetter, Lindgren and Rootzén [LLR83], [EKM97] and McNeil [McN98a]. See McNeil [McN98b] for a non-technical example pertaining to the block maxima method and the market crash of 1987.

6.4 Peaks Over Threshold Method

The more modern approach to modelling extreme events is to attempt to focus not only the largest (maximum) events, but on all events greater than some large preset threshold. This is referred to as *peaks over threshold* (POT) modelling. We will discuss two approaches to POT modelling currently found in the literature. The first is a semi-parametric approach based on a Hill type estimator of the tail index ([BTV96], Danielsson and de Vries [DdV97] and [DdV00], and Mills [Mil99]). The second approach is a fully parametric approach based on the generalized Pareto distribution ([EKM97], McNeil [MS97], and Embrechts, Resnick and Samorodnitsky [ERS99]).

6.4.1 Semiparametric Approach

Recall that F_X is in the maximum domain of attraction of the Fréchet distribution if and only if $\bar{F}_X(x) = x^{-\alpha}L(x)$ for some slowly varying function L . Suppose F_X is the distribution function of a loss distribution over some time horizon, where we would like to calculate a quantile based risk

measure such as VaR . Assume for simplicity that the distribution of large losses is of Pareto type

$$\mathbb{P}(X > x) = cx^{-\alpha}, \quad \alpha > 0, \quad x > x_0. \quad (43)$$

The semi-parametric approach uses a Hill type estimator for α and order statistics of historical data to invert and solve for VaR .

We first focus on VaR . Let $X^{(1)} \geq X^{(2)} \geq \dots \geq X^{(n)}$ be the order statistics of an historical sample of losses of size n , assumed *i.i.d.* with distribution F_X . If X is of Pareto type in the tail and $X^{(k+1)}$ is a high order statistic then for $x > X^{(k+1)}$,

$$\frac{\bar{F}_X(x)}{\bar{F}_X(X^{(k+1)})} = \left(\frac{x}{X^{(k+1)}} \right)^{-\alpha}.$$

The empirical distribution function estimator $\hat{F}_X(X^{(k+1)}) = k/n$ suggests the following estimator of F_X in the upper tail,

$$\hat{F}_X(x) = 1 - \frac{k}{n} \left(\frac{x}{X^{(k+1)}} \right)^{-\hat{\alpha}} \quad \text{for } x > X^{(k+1)}.$$

By inverting this relation, one can express x in terms of $\hat{F}_X(x)$, so that fixing $q = \hat{F}_X(x)$ one gets⁴⁸ $x = \widehat{VaR}_q(X)$. The value of q should be large, namely, $q = \hat{F}_X(x) > \hat{F}_X(X^{(k+1)}) = 1 - k/n$. This yields

$$\widehat{VaR}_q(X) = X^{(k+1)} \left(\frac{n}{k}(1 - q) \right)^{-1/\hat{\alpha}}. \quad (44)$$

We obtained an estimator for VaR but it depends on k through $X^{(k+1)}$, on the sample size n and $\hat{\alpha}$. To estimate α , Hill [Hil75] proposed the following estimator $\hat{\alpha}^{(Hill)}$ which is also dependent on the order statistics and sample size:

$$\hat{\alpha}^{(Hill)} = \hat{\alpha}_{k,n}^{(Hill)} = \left(\frac{1}{k} \sum_{i=1}^k \ln X^{(i)} - \ln X^{(k+1)} \right)^{-1}. \quad (45)$$

The consistency and asymptotic normality properties of this $\hat{\alpha}^{(Hill)}$ estimator are known in the *i.i.d.* case and for certain stationary processes. There are however, many issues surrounding Hill-type estimators, see for example [BTV96], [EKM97] and Drees, de Haan and Resnick [DdHR00].

To obtain $VaR_q(X)$, one also needs to choose the threshold level $X^{(k+1)}$ or, equivalently, k . Daniels, de Haan, Peng and de Vries [DHPV01] provide an optimal choice for k by means of a two stage bootstrap method. Even in this case, however, optimal means merely minimizing the *asymptotic* mean squared error, which leaves the user uncertain as to how to proceed in the finite sample case. Traditionally the choice of k is done visually by constructing a Hill plot.

The Hill plot $\{(k, \hat{\alpha}_{k,n}^{(Hill)}) : k = 1, \dots, n - 1\}$ is a visual check for the optimal choice of k . The choice of k and therefore of $\hat{\alpha}_{k,n}^{(Hill)}$, is inferred from a stable region of the plot since in the Pareto case, where (43) holds, $\hat{\alpha}_{n-1,n}^{(Hill)}$ is the maximum likelihood estimator for α . In the more general case

$$1 - F(x) \sim x^{-\alpha} L(x), \quad x \rightarrow \infty, \quad \alpha > 0, \quad (46)$$

⁴⁸We write here VaR_q and not VaR_α since now α represents the heavy-tail exponent.

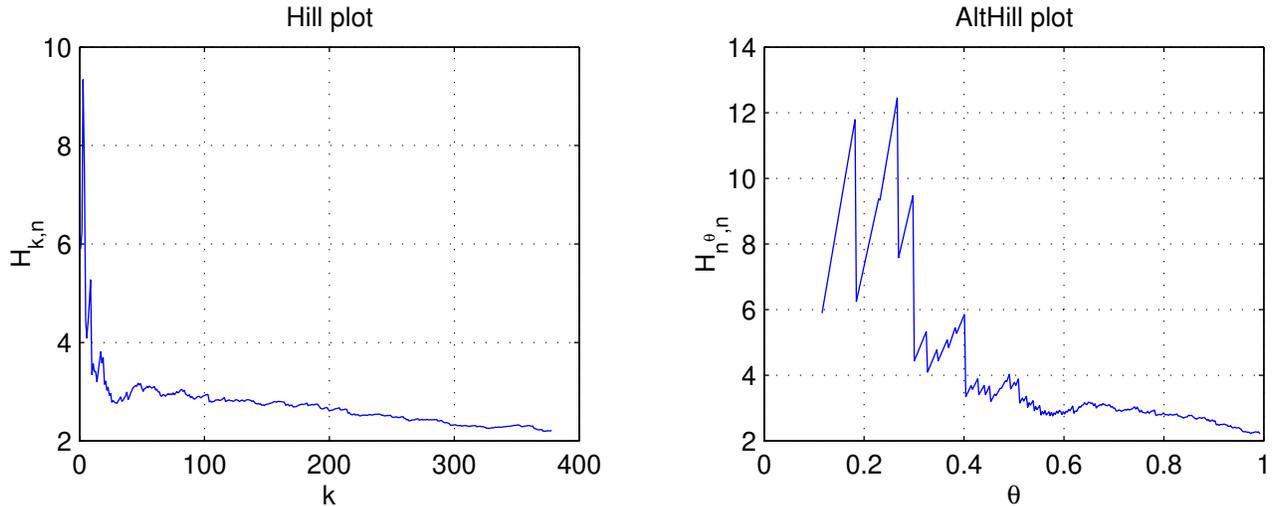


Figure 13: Hill plots for the *NASDAQ* data set. (Left) The Hill plot $\{(k, \hat{\alpha}_{k,n}^{(Hill)}) : k = 1, \dots, n-1\}$. (Right) The AltHill plot $\{(\theta, \hat{\alpha}_{\lceil n^\theta \rceil, n}^{(Hill)}) : 0 \leq \theta < 1\}$. The Hill plot is difficult to read, whereas the AltHill plot gives the user an estimate of $\hat{\alpha}^{AltHill} \approx 3$.

where L is a slowly varying function, the traditional Hill plot is often difficult to interpret. Resnick and Stărică [RS97] suggest an alternative plot, called an *AltHill* plot by plotting $\{(\theta, \hat{\alpha}_{\lceil n^\theta \rceil, n}^{(Hill)}) : 0 \leq \theta < 1\}$ where $\lceil n^\theta \rceil$ denotes the smallest integer greater than or equal to n^θ . This plot has the advantage of stretching the left hand side of the plot, which corresponds to smaller values of k , often making the choice of k easier. See Figure 13 for examples of the Hill and AltHill plots for the ordered negative returns $X^{(j)}$ for the *NASDAQ*.

6.4.2 Fully Parametric Approach

The fully parametric approach uses the generalized Pareto distribution (GPD) and the second fundamental theorem in EVT by Pickands, Balkema and de Haan. The GPD is a two-parameter distribution

$$G_{\xi, \beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} & : \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & : \xi = 0, \end{cases}$$

where an additional parameter $\beta > 0$ has been introduced. The support of $G_{\xi, \beta}(x)$ is $x \geq 0$ for $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ for $\xi < 0$. The distribution is heavy-tailed when $\xi > 0$. GPD distributions with $\beta = 1$ are displayed in Figure 14.

Definition 6.1 Let $X \sim F$ with right-end-point $x_F = \sup\{x \in \mathbb{R} \mid F(x) < 1\} \leq \infty$. For any high threshold $u < x_F$ define the excess distribution function

$$F_u(x) = \mathbb{P}(X - u \leq x \mid X > u) \quad \text{for } 0 \leq x < x_F - u. \quad (47)$$

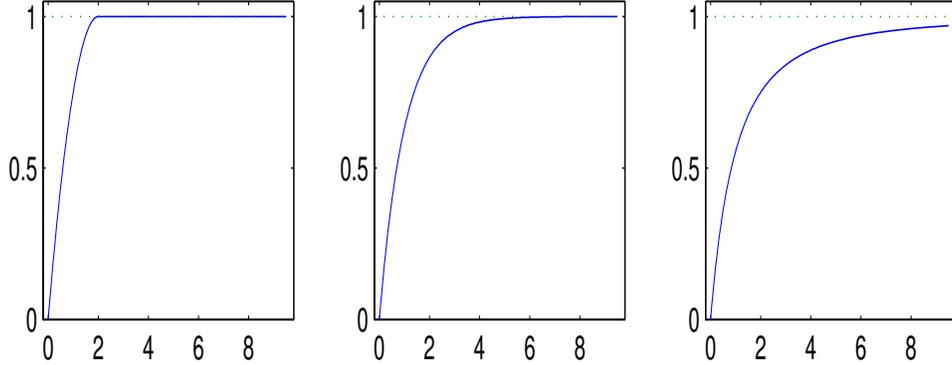


Figure 14: GPD distribution functions $G_{\xi, \beta}$, all with $\beta = 1$. (Left) $\xi = -0.5$, (Middle) $\xi = 0$, (Right) $\xi = 0.5$, which corresponds to a location adjusted Pareto distribution with $\alpha = 2$.

The mean excess function of X is then

$$e_X(u) = \mathbb{E}(X - u | X > u). \quad (48)$$

If X has exceeded the high level u , $F_u(x)$ measures the probability that it did not exceed it by more than x . Note that for $0 \leq x < x_F - u$, we may express $F_u(x)$ in terms of F ,

$$F_u(x) = \frac{F(u+x) - F(u)}{1 - F(u)},$$

and the mean excess function $e_X(u)$ may be expressed as a function of the excess distribution F_u as

$$e_X(u) = \int_0^{x_F - u} x dF_u(x).$$

The following theorem relates F_u to a GPD through the maximum domain of attraction of a GEV distribution. In fact, it completely characterizes the maximum domain of attraction of H_ξ .

Theorem 6.3 (Pickands (1975), Balkema and de Haan (1974)). *Let $X \sim F$. Then for every $\xi \in \mathbb{R}$, $X \in \text{MDA}(H_\xi)$ if and only if*

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0$$

for some positive function β .

This theorem says that the excess distribution F_u may be replaced by the GPD distribution G when u is very large. To see how it can be used, note that by (47) above, we may write

$$\bar{F}(x) = \bar{F}(u) \bar{F}_u(x - u) \quad (49)$$

for $x > u$. Assuming that u is sufficiently large, we may then approximate F_u by $G_{\xi, \beta(u)}$ and use the empirical estimator, for $\bar{F}(u)$,

$$\hat{\bar{F}}(u) = \frac{N_u}{n} \quad \text{where} \quad N_u = \sum_{i=1}^n 1_{\{X_i > u\}}$$

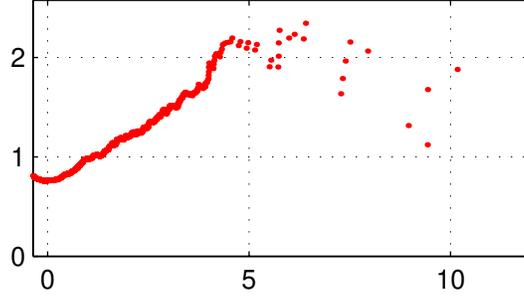


Figure 15: Sample mean excess plot $(u, \hat{e}_X(u))$ for *NASDAQ*.

and where n is the total number of observations. The upper tail of $F(x)$ may then be estimated by

$$\hat{F}(x) = 1 - \hat{F} = 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}} \quad \text{for all } x > u. \quad (50)$$

This way of doing things allows us to extrapolate beyond the available data which would not be possible had we chosen an empirical estimator for $F(x)$, $x > u$. We can therefore deal with potentially catastrophic events which have not yet occurred.

The parameters ξ and β of the GPD $G_{\xi, \beta(u)}$ may be estimated by using, for example, maximum likelihood once the threshold u has been chosen. The data points that are used in the maximum likelihood estimation are $X_{i_1} - u, \dots, X_{i_k} - u$ where X_{i_1}, \dots, X_{i_k} are the observations that exceed u . Again there is a bias-variance trade-off in the choice of u . To choose a value for u , a graphical tool known as the *mean excess plot* $(u, e_X(u))$ is often used.

The mean excess plot relies on the following theorem for generalized Pareto distributions.

Theorem 6.4 ([EKM97]) *Suppose X has GPD distribution with $\xi < 1$ and β . Then, for $u < x_F$,*

$$e_X(u) = \frac{\beta + \xi u}{1 - \xi}, \quad \beta + \xi u > 0.$$

The restriction $\xi < 1$ implies that the heavy-tailed distribution must have at least a finite mean.

If the threshold u is large enough so that F_u is approximately $G_{\xi, \beta}$ then, by Theorem 6.4, the plot $(u, e(u))$ is linear in u . How then is one to pick u ? The mean excess plot is a graphical tool for examining the relationship between the possible threshold u and the mean excess function $e_X(u)$ and checking the values of u where there is linearity. In practice it is not $e_X(u)$, but its sample version

$$\hat{e}_X(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n 1_{\{X_i > u\}}}$$

which is plotted against u . After using the mean excess plot to pick the upper threshold u one obtains an estimator of the tail of the distribution by applying (50). For the *NASDAQ* data, since linearity seems to start at relatively small values of u (Figure 15), we choose $u = 1.59$ which corresponds to the 95% of the empirical *NASDAQ* return distribution.

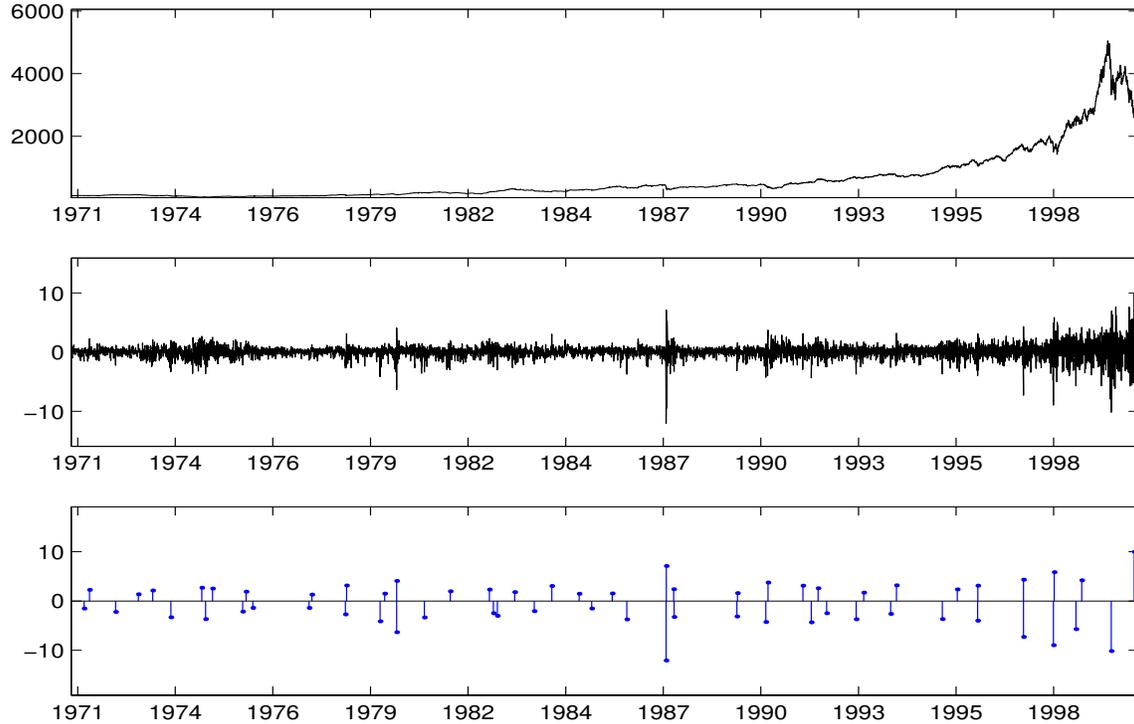


Figure 16: Time series of *NASDAQ* daily prices, (log) returns and annual maxima and minima daily returns given as a percent for the period February 1971 (when it was created) to February 2001. If P_t is the price (level) at time t , the returns are defined as $100 \ln(P_t/P_{t-1})$ and expressed as %. The crash of 1987 is clearly visible. The *NASDAQ* price level peaked in March of 2000.

To obtain $VaR_\alpha(X)$ for $VaR_\alpha(X) > u$, one simply inverts the tail estimator (50), which yields

$$\widehat{VaR}_\alpha(X) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right). \quad (51)$$

Since *expected shortfall* is a risk measure with better technical properties than VaR we would like to find an estimator for it which uses our GPD model of the tail. Recalling the definitions of the expected shortfall (22) and the mean excess function (48) we have that

$$S_\alpha(X) = VaR_\alpha(X) + e_X(VaR_\alpha(X)).$$

Since the excess distribution F_u is approximated by a GPD $G_{\xi, \beta(u)}$ with $\xi < 1$ then, applying Theorem 6.4, we get for $VaR_\alpha(X) > u$,

$$S_\alpha(X) = VaR_\alpha(X) + \frac{\beta + \xi (VaR_\alpha(X) - u)}{1 - \xi} = \frac{\beta + VaR_\alpha(X) - \xi u}{1 - \xi}.$$

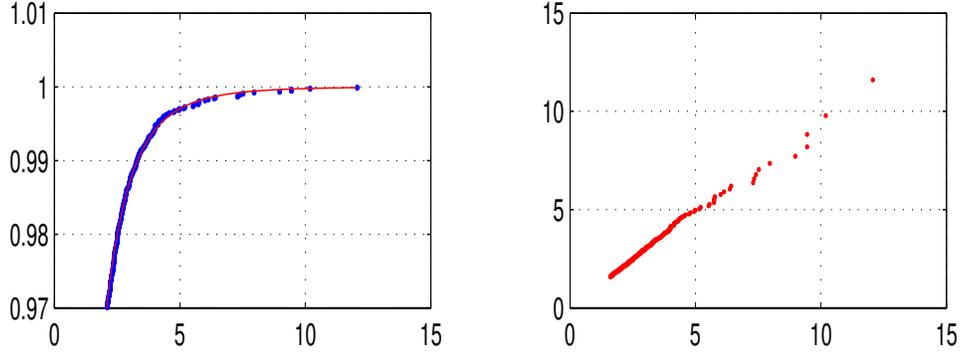


Figure 17: For the *NASDAQ* return data (as %), there were 379 exceedances above the high threshold $u = 1.59\%$. These are fitted with a GPD distribution $G_{\hat{\xi}, \hat{\beta}}$ with $\hat{\xi} = .189$ and $\hat{\beta} = .915$. (Left) The fitted GPD distribution (dark curve) and the empirical one (dotted curve). (Right) QQ-Plot of sample quantiles versus the quantiles of the fitted $G_{\hat{\xi}, \hat{\beta}}$ distribution.

This suggests the following estimator for expected shortfall,

$$\hat{S}_\alpha(X) = \frac{\hat{x}_\alpha}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}} \quad (52)$$

where $\hat{x}_\alpha = \widehat{VaR}_\alpha(X)$ may be obtained by using (51). As in the case of block maxima, confidence intervals for \widehat{VaR}_α and \hat{S}_α may be constructed using profile log-likelihood functions.

6.4.3 Numerical Illustration

To illustrate the usefulness of EVT in risk management, we consider the following example. Let X_1, \dots, X_n represent the daily negative returns of the *NASDAQ* index over most of its history from February 1971 to February 2001, which gives a time series of $n = 7570$ data points.

The price and return series are displayed in Figure 16. Let $X^{(1)} \geq \dots \geq X^{(n)}$ be the corresponding order statistics. Suppose the risk manager wants to obtain value at risk and expected shortfall estimates of the returns on the index at some high quantile. Assume that $\{X_i\}_{i=1}^n$ are *i.i.d.* so that Theorem 6.1 holds. Then, using Theorem 6.3, we model the tail of the excess distribution F_u by a GPD $G_{\xi, \beta}$ and use (49) to model the distribution $F(x)$ of the observations for all $x > u$. We use Theorem 6.4 and the sample mean excess plot, Figure 15, to pick the high threshold $u = 1.59\%$. This leaves us with $k = 379$ observations from which we estimate the parameters of the GPD by maximum likelihood. The estimates give $\xi = .189$ and $\beta = .915$. The model fit is checked by using a QQ plot displayed in Figure 17. Accepting the model, we go on to calculate the value at risk and expected shortfall for various high quantiles α by using (51) and (52). The results for the *NASDAQ* are plotted in Figure 18 (solid lines). If one had assumed that the observations were normally distributed (dashed lines), both the *VaR* and the expected shortfall would have been significantly underestimated for high quantiles.

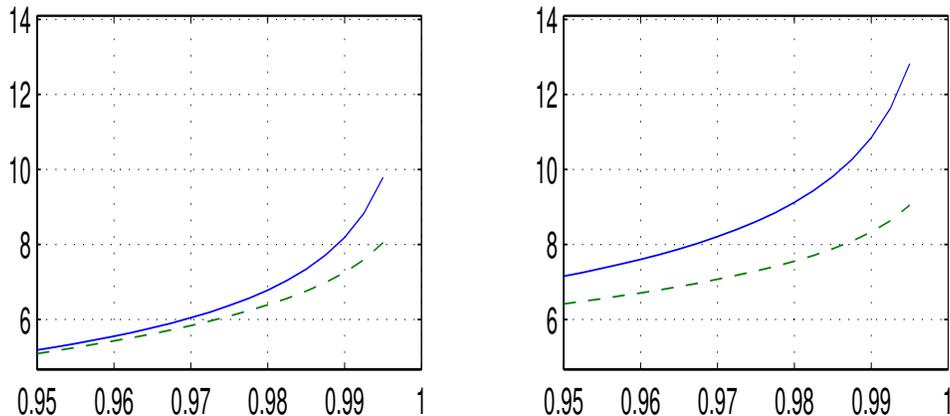


Figure 18: Risk estimates for *NASDAQ* in percent returns versus α . (Left) Value at risk VaR_α , for GPD (solid) and normal (dashed). (Right) Expected shortfall S_α , for GDP (solid) and normal (dashed). The parameters of the GPD are fitted by maximum likelihood using 30 years of data. The sample mean and volatility of the normal distribution are computed by (16) using the most recent year of daily observations.

For example, at the $\alpha = .99$ confidence level, $\widehat{VaR}_\alpha(X) = 6.59\%$ under the normal model versus $\widehat{VaR}_\alpha(X) = 8.19\%$ for the GPD model. For the expected shortfall, the difference is even more dramatic. For the normal model, $\widehat{S}_\alpha(X) = 7.09\%$ versus $\widehat{S}_\alpha(X) = 10.8\%$ for the GPD model. This is to be expected, since under the assumption of normality it may be shown ([EKM97]) that

$$\frac{S_\alpha}{VaR_\alpha} \longrightarrow 1 \quad \text{as } \alpha \rightarrow 1^-,$$

whereas for the GPD model

$$\frac{S_\alpha}{VaR_\alpha} \longrightarrow \frac{1}{1-\xi} \quad \text{as } \alpha \rightarrow 1^-.$$

These results indicate that for very high quantiles, the expected shortfall S_α and the value at risk VaR_α are comparable under normality, but for the GPD with $\xi < 1$, S_α tends to be larger than VaR_α .

6.4.4 A GARCH-EVT model for Risk

In order to invoke Theorems 6.1 and 6.3 in the numerical illustration above it was necessary to assume that the (negative) returns $\{X_t\}_{t \in \mathbb{Z}}$ were *i.i.d.* However, from inspection of Figures 16 and 19, it is apparent that this assumption is unrealistic. The time series of returns is characterized by periods of varying volatility, that is, the time series is *heteroscedastic*. The heteroscedasticity of the time series may cause problems for the estimation of the parameters of the GPD model since we would expect the high threshold u to be violated more often during periods of high volatility. Smith [Smi00] suggests using Bayesian techniques to model time-varying GPD parameters. In this section, we review a model proposed by McNeil and Frey [MF00] which extends the EVT methodology to models of financial time series that allow for stochastic volatility and apply this model to the *NASDAQ* data set.

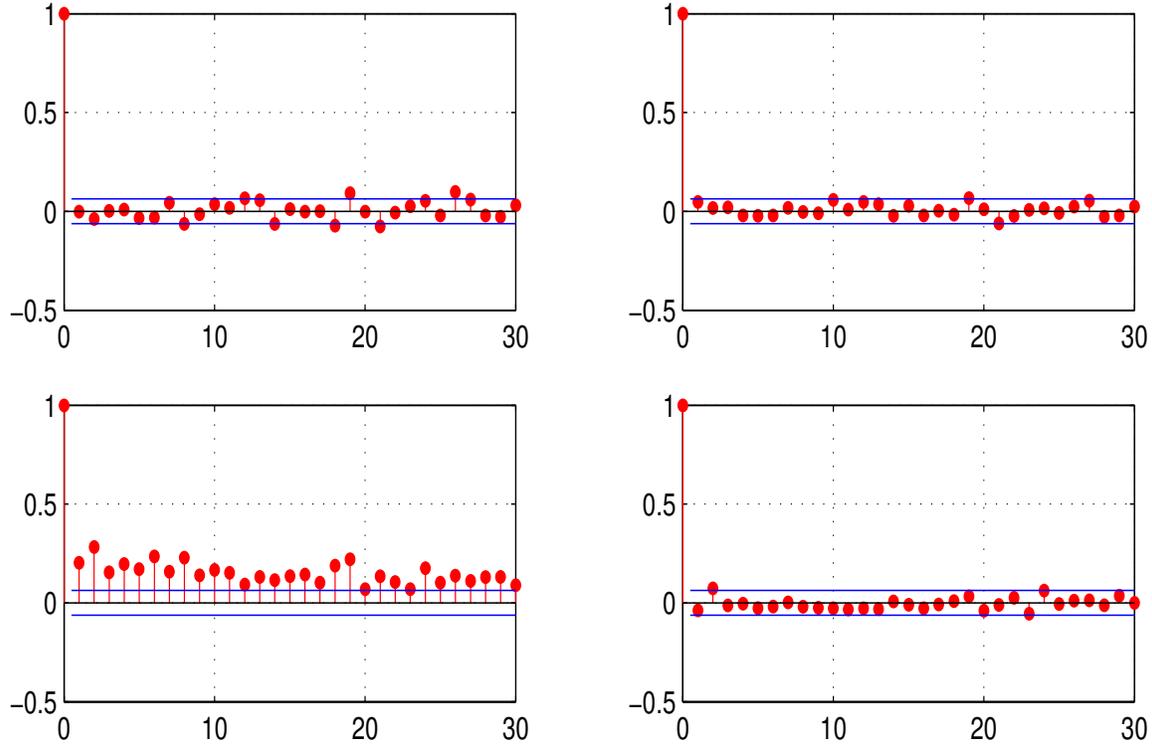


Figure 19: Sample auto correlation functions with lags on the abscissa and sample autocorrelation on the ordinate: Returns (top left), squared Returns (bottom left), GARCH Innovations (top right), squared GARCH Innovations (bottom right). The sample consists of 1000 daily returns for the *NASDAQ* ending Feb. 2001. Horizontal lines indicate the 95% confidence bands ($\pm 1.96/\sqrt{n}$) corresponding to Gaussian white noise.

Recall from Section 3.2.2 that the standard GARCH(1,1) model is given by⁴⁹

$$X_t = \sigma_t Z_t \text{ where } Z_t \sim F_Z \text{ i.i.d.}, \quad (53)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (54)$$

Since the time $t + 1$ volatility σ_{t+1} is known at time t we have that

$$\begin{aligned} VaR_\alpha(X_{t+1} | \mathcal{F}_t) &:= \inf \{x \in \mathbb{R} | F_{X_{t+1} | \mathcal{F}_t}(x) \geq \alpha\} \\ &= \sigma_{t+1} z_\alpha, \end{aligned} \quad (55)$$

where $z_\alpha = F_Z^{-1}(\alpha)$. The same argument shows that the conditional expected shortfall

$$S_\alpha(X_{t+1} | \mathcal{F}_t) := \mathbb{E}(X_{t+1} | X_{t+1} > VaR_\alpha(X_{t+1} | \mathcal{F}_t), \mathcal{F}_t) = \sigma_{t+1} \mathbb{E}(Z | Z > z_\alpha).$$

⁴⁹Since the *NASDAQ* series appears to have a zero conditional mean we do not set $X_t = \mu_t + \sigma_t Z_t$ and model the mean μ_t , for example as an AR(1) process $\mu_t = \phi X_{t-1}$.

Traditionally the innovation distribution F_Z is assumed normal. Figures 6 and 20 show that this assumption may still underestimate the tails of the loss portion of the distribution. McNeil and Frey propose a two step procedure to estimate VaR and expected shortfall of the conditional distribution. First they use a GARCH(1,1) model for the volatility of the (negative) return series $\{X_t\}$. This gives a series of model implied innovations $Z_t = X_t/\sigma_t$. Second, EVT is used to model the tails of the distribution of these innovations. This approach has the obvious benefit that the resulting innovations Z_t are much closer to satisfying the requirements of Theorems 6.1 and 6.3 than is the original series. We illustrate the methodology with an example using the *NASDAQ* data.

1. Let $(x_{t-n+1}, \dots, x_{t-1}, x_t)$ be n daily negative returns of the *NASDAQ*. We take⁵⁰ $n = 1000$ and use *pseudo-maximum-likelihood* (PML) to estimate the model parameters $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$ in (54) under the assumption⁵¹ that F_Z is normal in (53). The parameter vector $\hat{\theta}$ depends on the true distribution of $(X_{t-n+1}, \dots, X_{t-1}, X_t)$, which is assumed stationary, and on the distribution F_Z used to compute the likelihood function⁵². When we assume F_Z is normal we fit a model whose distributional assumptions we do not believe. Under standard regularity conditions this is justified since $\hat{\theta}$ is a consistent estimator of θ (in fact, asymptotically normal) even if F_Z is non normal. See Gouriéroux [Gou97] and references therein for details.
2. The model innovations $(z_{t-n+1}, \dots, z_{t-1}, z_t) = (x_{t-n+1}/\hat{\sigma}_{t-n+1}, \dots, x_{t-1}/\hat{\sigma}_{t-1}, x_t/\hat{\sigma}_t)$ are now calculated. If the model is tenable, these innovations should be *i.i.d.* Figure 19 shows that while the *i.i.d.* assumption is not realistic for the series of returns, it is defensible for the series of innovations⁵³. While the returns appear uncorrelated, their squares clearly are not, and hence the returns are dependent. The GARCH innovations and their squares appear uncorrelated. The *i.i.d.* assumption is therefore more tenable.
3. Examination of the QQ plot of the innovations in Figure 20 reveals that the loss tail is heavier than that of the normal. Therefore the EVT tools of Section 6.4.2 are now applied to the innovations $(z_{t-n+1}, \dots, z_{t-1}, z_t)$. Let $z^{(1)} \leq \dots \leq z^{(n)}$ be the order statistics of the innovation sample. We choose the threshold $u = 1.79$, again corresponding to the 95% of the empirical distribution of innovations, which leaves $k = 50$ observations $(z^{(n-k+1)}, \dots, z^{(n)})$, from which to estimate the GPD parameters by maximum likelihood. The estimates give $\xi = .323$ and $\beta = .364$.

Observe that $\xi = .323$ corresponds to a heavier tail than $\xi = .189$ which we found in Section 6.4.3. We are fitting here, however, over a particularly volatile period of 1000 days of the *NASDAQ* ending Feb. 2001, whereas in Section 6.4.3, we considered nearly 30 years worth of returns where for the majority of the time the *NASDAQ* was significantly less volatile (see Figure 16).

Since the model is assumed stationary, we could, in principle, use the estimated GARCH parameters to compute $\hat{\sigma}_{t+1} | \mathcal{F}_t$ using (54) for t beyond Feb. 2001. Using z_α corresponding to the GPD distribution $G_{\xi, \beta}$, we would obtain, by using (55), $\widehat{VaR}_\alpha(X_{t+1} | \mathcal{F}_t)$ for t beyond Feb. 2001. In practice, however,

⁵⁰We keep the sample size moderate in order to avoid the *IGARCH* effect, that is $\alpha_1 + \beta_1 = 1$, corresponding to non-stationarity. See [MS00] for details.

⁵¹The term *pseudo* refers to the fact that one is not maximizing the true likelihood.

⁵²The condition $\alpha_1 + \beta_1 < 1$ is sufficient for stationarity of the GARCH model. We found $\hat{\alpha}_0 = .080$, $\hat{\alpha}_1 = .181$ and $\hat{\beta}_1 = .811$. However, as indicated in the sequel, the GARCH model is constantly updated, and hence is never used on an infinite horizon.

⁵³Ljung-Box tests also found no evidence against the *i.i.d.* assumption for the innovations.

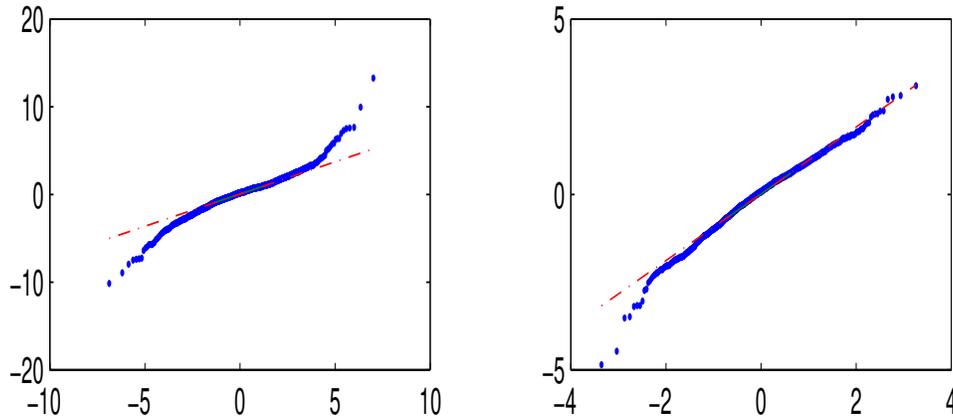


Figure 20: QQ plots versus the normal for returns (left) and innovations (right) in Figure 19. Notice that the lower (loss) tail of the innovations are still heavier than the normal distribution.

stationarity is not always assured and in any case one wants to use the most recent data available in order to calibrate the model.

In order to backtest the methodology we use the most recent 500 days in our *NASDAQ* data set. For each day, $t + 1$, in this data set we use the previous $n = 1000$ days (negative) returns $(X_{t-n+1}, \dots, X_{t-1}, X_t)$ to calibrate the model and estimate $VaR_\alpha(X_{t+1} | \mathcal{F}_t)$ for $\alpha = .95$ and $\alpha = .99$ using the steps above. We compare $\widehat{VaR}_\alpha(X_{t+1} | \mathcal{F}_t)$ with the actual loss x_{t+1} . A violation, at the α level, is said to occur whenever $x_{t+1} > \widehat{VaR}_\alpha(X_{t+1} | \mathcal{F}_t)$. Results for the period ending Feb. 2001 are given in Figure 21.

7 Stable Paretian Models

The works of Mandelbrot [Man63] and Fama [Fam65] introduced the use of stable distributions to finance. The *excessively peaked* and *heavy-tailed* nature of the return distribution led the authors to reject the standard hypothesis of normally distributed returns in favor of the stable distribution. Since this time, the stable distribution has been used to model both the unconditional, and conditional return distributions. In addition, portfolio theories and market equilibrium models have been constructed using it. For an in depth introduction to the general properties of stable distributions see Samorodnitsky and Taqqu [ST94] and the upcoming text Nolan [Nol01]. A major reference for applications in finance is Rachev and Mittnik [RM00].

In Definition 3.2, the stable distribution $S_\alpha(\sigma, \beta, \mu)$ is defined as the limiting distribution of the sum of *i.i.d.* random variables. Like the normal distribution, stable distributions are closed under addition, and are often defined by this property. Recall that if $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Similarly, for stable random variables, if $X_1 \sim S_\alpha(\sigma_1, \beta_1, \mu_1)$ and $X_2 \sim S_\alpha(\sigma_2, \beta_2, \mu_2)$ are independent, then $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$ where

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \mu = \mu_1 + \mu_2.$$

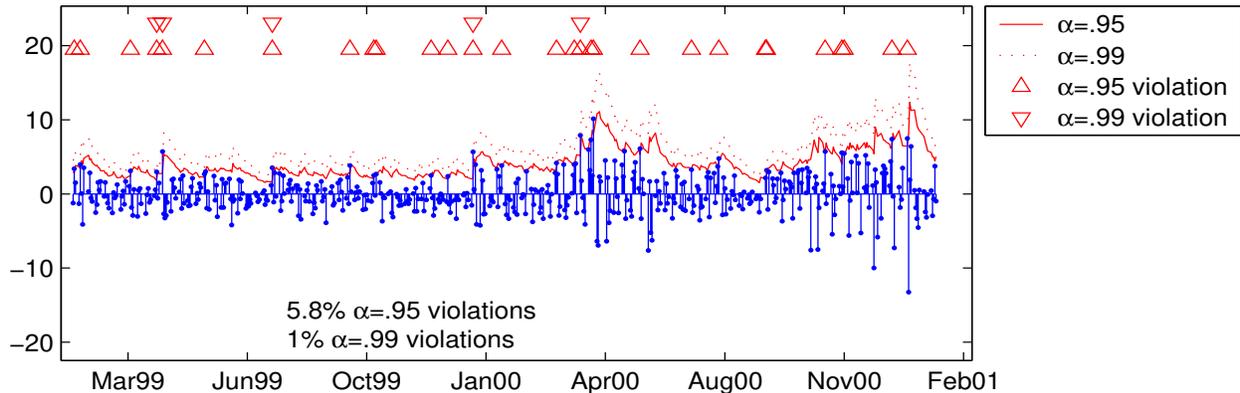


Figure 21: Backtest results for the GARCH-EVT methodology of McNeil and Frey. Under the assumption that the model correctly estimates the conditional quantiles we expect violations 5% and 1% of the time for $\alpha = .95$ and $\alpha = .99$ respectively. VaR for $\alpha = .95$ and $\alpha = .99$ are given by the solid and dotted lines respectively. We obtain 5.8% violations of the $\alpha = .95$ level and 1% violations of the $\alpha = .99$ level.

It is in this sense that the stable distribution is a natural heavy-tailed alternative to the normal distribution. However, a common criticism of the stable distribution is that their tails are too heavy. One has $\mathbb{P}(X > x) \sim c_\alpha x^{-\alpha}$ as $x \rightarrow \infty$. For $0 < \alpha < 2$, this implies that $\mathbb{E}|X|^p < \infty$ if $0 < p < \alpha$. In particular, $\mathbb{E}X^2 = \infty$, that is, all non-Gaussian stable distributions have infinite variance.

The stable distributions can be defined and parameterized in different ways. One way to specify a stable distribution is through its characteristic function. This is helpful since in general there exists no closed form for the probability density function⁵⁴, which historically, has been an impediment to their widespread use. Today, however, there are efficient computer programs to evaluate their densities using fast Fourier transform methods ([RM00] and [Nol01]).

Definition 7.1 A random variable X is said to have a stable distribution if there are parameters $\alpha \in (0, 2]$, $\sigma \in [0, \infty)$, $\beta \in [-1, 1]$ and $\mu \in \mathbb{R}$ such that its characteristic function has the following form:

$$\Psi_X(t) = \begin{cases} \exp \left\{ -\sigma^\alpha |t|^\alpha \left(1 - i\beta (\text{sign } t) \tan \frac{\pi\alpha}{2} + i\mu t \right) \right\} & \text{for } \alpha \neq 1, \\ \exp \left\{ -\sigma |t| \left(1 + i\beta \frac{2}{\pi} (\text{sign } t) \ln |t| \right) + i\mu t \right\} & \text{for } \alpha = 1. \end{cases} \quad (56)$$

If both the skewness and location parameters β and μ are zero, X is said to be symmetric stable, which is denoted $X \sim S\alpha S$, and its characteristic function takes the simple form

$$\Psi_X(t) = e^{-\sigma^\alpha |t|^\alpha}.$$

If $X \sim S\alpha S$, then it is characterized completely by its index of stability α and its scale parameter σ . If $\alpha = 2$, the Gaussian case, then the scale parameter is $\sigma = \sqrt{\frac{1}{2}\text{Var}(X)}$.

⁵⁴The exceptions to this rule are the distributions $S_2(\sigma, 0, \mu)$, $S_1(\sigma, 0, \mu)$, and $S_{1/2}(\sigma, 1, \mu)$ which correspond to the Gaussian, Cauchy and Lévy distributions respectively.

7.1 Stable Portfolio Theory

In Section 2.2 we introduced the mean-variance portfolio theory of Markowitz. The model assumed that the distribution of asset returns is multivariate normal, and provides efficient portfolios, that is, portfolios with maximum expected return for a given level of risk, where risk is measured by the variance of the portfolio. It is possible to extend the ideas of portfolio theory to the case where asset returns have a multivariate stable distribution, even though, variances are now infinite. We need first to define a stable random vector and specify its characteristic function.

Definition 7.2 *The random vector $\mathbf{X} = (X_1, \dots, X_n)$ is said to be a stable random vector in \mathbb{R}^n if for any $a, b > 0$ there exists $c > 0$ and $\mathbf{d} \in \mathbb{R}^n$ such that*

$$a\mathbf{X}_1 + b\mathbf{X}_2 \stackrel{d}{=} c\mathbf{X} + \mathbf{d} \quad (57)$$

where $\mathbf{X}_j, j = 1, 2$ are independent copies of \mathbf{X} .

The constants in (57) are related by $c^\alpha = a^\alpha + b^\alpha$, where $\alpha \in (0, 2]$ is the index of stability. Setting $n = 1$ in (57) yields one of the alternate definitions of a stable random variable alluded to earlier. In the case of a stable random vector, the scale and skewness parameters σ and β are replaced by a finite measure $\Gamma_{\mathbf{X}}$ on the unit hypersphere in \mathbb{R}^n . For convenience here, let (\cdot, \cdot) denote the inner product so that $(\mathbf{t}, \mathbf{s}) = \sum_{i=1}^n t_i s_i$ ⁵⁵.

Theorem 7.1 *Let $0 < \alpha < 2$. Then $\mathbf{X} = (X_1, \dots, X_n)$ is a stable random vector with index of stability α if and only if there exists a finite measure $\Gamma_{\mathbf{X}}$ on the unit hypersphere $S_n = \{\mathbf{s} \in \mathbb{R}^n \mid \|\mathbf{s}\| = 1\}$ and a vector $\boldsymbol{\mu} \in \mathbb{R}^n$ such that*

$$\Psi_\alpha(\mathbf{t}) = \begin{cases} \exp \left\{ - \int_{S_n} |(\mathbf{t}, \mathbf{s})|^\alpha (1 - i \operatorname{sign}((\mathbf{t}, \mathbf{s})) \tan \frac{\pi\alpha}{2}) \Gamma_{\mathbf{X}}(d\mathbf{s}) + i(\mathbf{t}, \boldsymbol{\mu}) \right\} & \text{for } \alpha \neq 1, \\ \exp \left\{ - \int_{S_n} |(\mathbf{t}, \mathbf{s})| (1 + i \frac{2}{\pi} \operatorname{sign}((\mathbf{t}, \mathbf{s})) \ln |(\mathbf{t}, \mathbf{s})|) \Gamma_{\mathbf{X}}(d\mathbf{s}) + i(\mathbf{t}, \boldsymbol{\mu}) \right\} & \text{for } \alpha = 1. \end{cases} \quad (58)$$

The pair $(\Gamma_{\mathbf{X}}, \boldsymbol{\mu})$ is unique.

The measure $\Gamma_{\mathbf{X}}$ is called the *spectral measure* of the stable random vector \mathbf{X} and specifies the dependence structure. If \mathbf{X} is $S\alpha S$ in \mathbb{R}^n , then the characteristic function takes the simple form

$$\Psi_\alpha(\mathbf{t}) = \exp \left\{ - \int_{S_n} |(\mathbf{t}, \mathbf{s})|^\alpha \Gamma_{\mathbf{X}}(d\mathbf{s}) \right\},$$

where Γ is the unique symmetric spectral measure. The expression in (58) for the characteristic function is also valid for the normal case $\alpha = 2$. When $\alpha = 2$, it reduces to $\Psi_2(\mathbf{t}) = \exp \left\{ - \int_{S_n} |(\mathbf{t}, \mathbf{s})|^2 \Gamma_{\mathbf{X}}(d\mathbf{s}) \right\}$ but in this case $\Gamma_{\mathbf{X}}$ is no longer unique. To get a feeling for $\Gamma_{\mathbf{X}}$, suppose $\mathbf{X} = (X_1, X_2)$ and that the distribution is Gaussian. Then

$$\begin{aligned} \int_{S_2} |(\mathbf{t}, \mathbf{s})|^2 \Gamma_{(X_1, X_2)}(d\mathbf{s}) &= \int_{S_2} |(t_1 s_1 + t_2 s_2)|^2 \Gamma_{(X_1, X_2)}(d\mathbf{s}) \\ &= t_1^2 \sigma_1^2 + 2t_1 t_2 \sigma_{1,2} + t_2^2 \sigma_2^2 \end{aligned}$$

⁵⁵Previously we wrote $\mathbf{t}^T \mathbf{s}$ instead of (\mathbf{t}, \mathbf{s}) .

where

$$\sigma_i^2 = \int_{S_2} s_i^2 \Gamma_{(X_1, X_2)}(d\mathbf{s}), \quad i = 1, 2 \quad \text{and} \quad \sigma_{1,2} = \int_{S_2} s_1 s_2 \Gamma_{(X_1, X_2)}(d\mathbf{s}),$$

and where integration over the circle S_2 means integration on $\{\mathbf{s} = (s_1, s_2) \mid s_1^2 + s_2^2 = 1\}$. One recognizes the normal characteristic function with $\text{Var}X_1 = 2\sigma_1^2$, $\text{Var}X_2 = 2\sigma_2^2$ and $\text{Cov}(X_1, X_2) = 2\sigma_{1,2}$. Since different choices of $\Gamma_{(X_1, X_2)}$ can yield the same values for σ_1^2 , σ_2^2 and $\sigma_{1,2}$, the choice of $\Gamma_{\mathbf{X}}$ is not unique in the Gaussian case.

As in the case of a normal random vector, if \mathbf{X} is multivariate stable with index of stability $0 < \alpha < 2$, then all linear combinations of the components of \mathbf{X} are stable with the same α . So, if \mathbf{X} is a stable random vector in \mathbb{R}^n , and $\mathbf{w} \in \mathbb{R}^n$, we know that $Y = (\mathbf{w}, \mathbf{X}) = \sum_{i=1}^n w_i X_i$ is $S_\alpha(\sigma_Y, \beta_Y, \mu_Y)$. Using the characteristic function (58), it can be shown (see [ST94], Example 2.3.4), that

$$\sigma_Y = \left(\int_{S_n} |(\mathbf{w}, \mathbf{s})|^\alpha \Gamma_{\mathbf{X}}(d\mathbf{s}) \right)^{1/\alpha}, \quad (59)$$

$$\beta_Y = \frac{\int_{S_n} |(\mathbf{w}, \mathbf{s})|^\alpha \text{sign}(\mathbf{w}, \mathbf{s}) \Gamma_{\mathbf{X}}(d\mathbf{s})}{\int_{S_n} |(\mathbf{w}, \mathbf{s})|^\alpha \Gamma_{\mathbf{X}}(d\mathbf{s})}, \quad (60)$$

$$\mu_Y = \begin{cases} (\mathbf{w}, \boldsymbol{\mu}) & \text{for } \alpha \neq 1, \\ (\mathbf{w}, \boldsymbol{\mu}) - \frac{2}{\pi} \int_{S_n} (\mathbf{w}, \mathbf{s}) \ln |(\mathbf{w}, \mathbf{s})| \Gamma_{\mathbf{X}}(d\mathbf{s}) & \text{for } \alpha = 1. \end{cases} \quad (61)$$

In the mean-variance portfolio theory, the risk to be minimized for any level of expected return is given by the portfolios' variance. If the asset returns are assumed multivariate stable with index of stability $0 < \alpha < 2$ then the variance is infinite and cannot be used. In the stable portfolio theory, it is assumed that $1 < \alpha < 2$, $\mathbb{E}\mathbf{X} = \boldsymbol{\mu}$ and that $\mathbf{X} - \boldsymbol{\mu} \sim S_\alpha S$. Let \mathbf{w} be the vector of weights for the risky portfolio $X_p = (\mathbf{w}, \mathbf{X})$. Given the relationship between the scale parameter and the variance in the Gaussian case (that is, stable with $\alpha = 2$), it is natural to use the scale parameter σ_{X_p} of the resulting stable distribution instead of the standard deviation. It is given by (59). This brings us to the corresponding stable portfolio problem:

$$\begin{aligned} \min_{\mathbf{w}} \sigma_{X_p} &= \left(\int_{S_n} |(\mathbf{w}, \mathbf{s})|^\alpha \Gamma_{\mathbf{X}}(d\mathbf{s}) \right)^{1/\alpha}, \\ \text{such that } (\mathbf{w}, \boldsymbol{\mu}) &\geq a, \\ (\mathbf{w}, \mathbf{e}) &= 1. \end{aligned} \quad (62)$$

The risk measure $\sigma_{X_p} = \sigma_{(\mathbf{w}, \mathbf{X})}$ is a convex function of \mathbf{w} and the problem is generally solved using sequential quadratic programming. See Belkacem [Bel97] and [RM00] and references therein for details of the procedure and on the estimation of the index of stability, spectral measure and scale parameters. If a risk free asset is included in the asset universe, then we end up with a maximization problem similar to (2) in Section 2.2, but where the risk measure is the scale parameter σ_{X_p} of the risky portfolio.

7.2 Stable Asset Pricing

Since there exists a portfolio theory under the assumption of a multivariate stable distribution of asset returns ($1 < \alpha < 2$), it is natural to ask whether there exists an analogous CAPM. The answer is

positive, and it was first introduced by Fama [Fam70]. For recent descriptions of the stable CAPM see Belkacem, Lévy Véhel and Walter [BVW96] and, of course, Rachev and Mittnik [RM00].

The assumptions behind the stable CAPM are the same as in the Gaussian case in Section 2.3 with the assumption of joint normality of asset returns replaced by that of jointly stable asset returns with index of stability $\alpha \in (1, 2)$. That is, we assume $\mathbb{E}\mathbf{X} = \boldsymbol{\mu}$ and that $\mathbf{X} - \boldsymbol{\mu} \sim S\alpha S$. Recall from the traditional CAPM and Equations (3) and (4), that the expected premium of holding the risky asset i over the riskless asset is proportional to the expected premium of holding the market portfolio over the riskless asset. The constant of proportionality was the risky assets *beta* given by (4). In the stable CAPM, we require an alternative measure of dependence since covariances do not exist. Naturally, the scale parameter σ replaces the standard deviation.

The *covariation* is a natural alternative to the covariance in the stable case when $1 < \alpha < 2$. This measure possesses many, but not all, of the useful properties of covariance in the Gaussian case. We define and present several of the properties of covariation. Details may be found in [ST94] and [RM00].

Definition 7.3 *Let X_1 and X_2 be jointly $S\alpha S$ with $1 < \alpha \leq 2$ and let $\Gamma_{(X_1, X_2)}$ be the spectral measure of the random vector (X_1, X_2) . The covariation of X_1 on X_2 is given by*

$$[X_1, X_2]_\alpha = \int_{S_2} s_1 s_2^{\langle \alpha - 1 \rangle} \Gamma_{(X_1, X_2)}(ds) \quad (63)$$

where $s^{\langle p \rangle}$ denotes the signed power $s^{\langle p \rangle} = |s|^p (\text{sign } s)$.

In the Gaussian case $\alpha = 2$ it reduces to

$$[X_1, X_2]_2 = \frac{1}{2} \text{Cov}(X_1, X_2). \quad (64)$$

Note, however, that whereas in the Gaussian case the dependence structure is fully characterized by the covariance, in the stable one needs to use $\Gamma_{\mathbf{X}}$, and the covariation does *not* fully characterize the dependence structure. We now derive the stable CAPM under the preceding assumptions, following [BVW96].

Consider a portfolio of a riskless asset with rate of return r and a risky asset X_i with weights w and $1 - w$ respectively. The expected rate of return of the portfolio $X_p = wr + (1 - w)X_i$ is then $\mathbb{E}X_p = wr + (1 - w)\mathbb{E}X_i$, and its risk, as given by its scale parameter, is $\sigma_p = (1 - w)\sigma_i$ ⁵⁶. The risk-return trade-off is then given by

$$\mathbb{E}X_p = r + \frac{\mathbb{E}X_i - r}{\sigma_i} \sigma_p \quad (65)$$

after setting $w = 1 - \sigma_p/\sigma_i$. Under the assumptions of CAPM, investors have homogeneous beliefs, that is, they all agree on the multivariate stable parameters. This means that all investors hold the market portfolio (as in Section 2.3) as their risky asset and the risk-return trade-off (65) becomes

$$\mathbb{E}X_p = r + \frac{\mathbb{E}X_M - r}{\sigma_M} \sigma_p. \quad (66)$$

where X_M and σ_M are the rate of return and scale parameter respectively of the market.

⁵⁶Note that if $X \sim S_\alpha(\sigma, \beta, \mu)$ then $aX + b \sim S_\alpha(|a|\sigma, \text{sign}(a)\beta, a\mu + b)$ if $1 < \alpha < 2$.

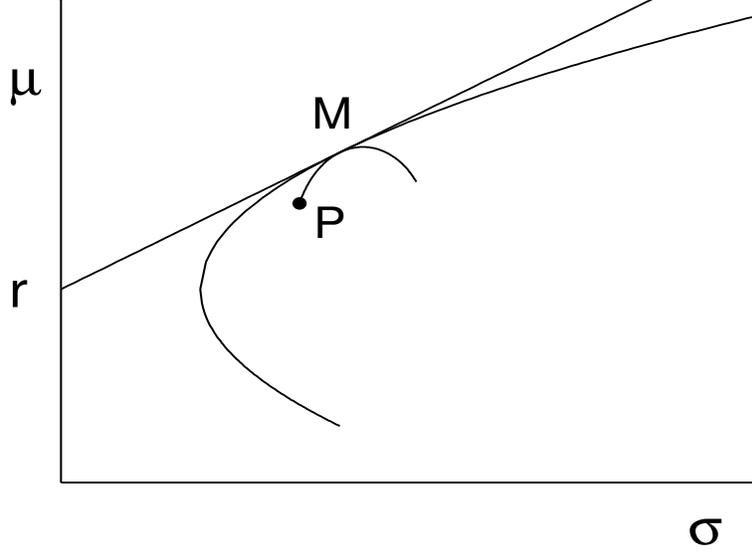


Figure 22: The stable efficient frontier. The portfolio $X_p = wX_i + (1 - w)X_M$ is suboptimal, and hence must be dominated by the efficient frontier.

Now consider the suboptimal portfolio $X_p = wX_i + (1 - w)X_M$ obtained by adding to the market portfolio a certain position in asset i (the portfolio is optimal if $w = 0$). Since $\mathbf{X} - \boldsymbol{\mu} \sim S\alpha S$ we know that $X_i - \mu_i$ and $X_M - \mu_M$ are jointly $S\alpha S$. By properties of symmetric stable random vectors this means that $X_p \sim S_\alpha(\sigma_p, 0, \mu_p)$, where the scale and location parameters are given by (59) and (61), that is

$$\sigma_p^\alpha = \int_{S_2} |ws_1 + (1 - w)s_2|^\alpha \Gamma_{(X_i, X_p)}(ds_1, ds_2), \quad (67)$$

$$\mu_p = \mathbb{E}X_p = w\mu_i + (1 - w)\mu_M, \quad (68)$$

respectively. Differentiating with respect to w gives

$$\frac{\partial \mu_p}{\partial w} = \mu_i - \mu_M, \quad (69)$$

$$\frac{\partial \sigma_p}{\partial w} = \frac{1}{\alpha \sigma_p^{\alpha-1}} \frac{\partial \sigma_p^\alpha}{\partial w} = \frac{1}{\sigma_p^{\alpha-1}} \int_{S_2} (s_1 - s_2)(ws_1 + (1 - w)s_2)^{\alpha-1} \Gamma_{(X_i, X_p)}(ds_1, ds_2). \quad (70)$$

So evaluating (69) and (70) at $w = 0$ and using Definition 7.3 we get

$$\left. \frac{\partial \mu_p}{\partial \sigma_p} \right|_{w=0} = \left. \frac{\partial \mu_p}{\partial w} / \frac{\partial \sigma_p}{\partial w} \right|_{w=0} = \frac{\sigma_M^{\alpha-1}(\mu_i - \mu_M)}{[X_i, X_M]_\alpha - \sigma_M^\alpha}, \quad (71)$$

since at $w = 0$ the portfolio X_p becomes X_M and σ_p becomes σ_M . Moreover, in market equilibrium the trade-off between risk and return is given by (66), so that the slope $\partial \mu_p / \partial \sigma_p$ at $w = 0$ is given by $(\mu_M - r) / \sigma_M$ (see Figure 22). Hence

$$\frac{\mu_M - r}{\sigma_M} = \frac{\sigma_M^{\alpha-1}(\mu_i - \mu_M)}{[X_i, X_M]_\alpha - \sigma_M^\alpha}. \quad (72)$$

This may be rewritten in the familiar CAPM form (3) as

$$\mathbb{E}(X_i - r) = \beta_i \mathbb{E}(X_M - r),$$

where now, in the stable case,

$$\beta_i = \frac{[X_i, X_M]_\alpha}{\sigma_M^\alpha}. \quad (73)$$

Note that if we assume Gaussian returns, then $\mathbf{X} - \boldsymbol{\mu} \sim S\alpha S$ with $\alpha = 2$, and by using (64), we recover

$$\beta_i = \frac{\text{Cov}(X_i, X_M)}{\text{Var}(X_M)},$$

that is, the traditional CAPM result.

Acknowledgments

We would like to thank Paul Embrechts and Filip Lindskog for many valuable comments which led to an improved exposition of this material. This research was partially supported by the NSF Grant ANI-9805623 at Boston University.

References

- [ADEH97] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Thinking coherently. *RISK*, 10(11), 1997.
- [ADEH99] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- [Alb97] C. Albanese. Credit exposure. Diversification risk and coherent VaR. Preprint, Department of Mathematics, University of Toronto, 1997.
- [BCK92] T. Bollerslev, R.Y. Chou, and K.F. Kroner. ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52:307–327, 1992.
- [Bel97] L. Belkacem. How to select optimal portfolio in α -stable markets. Preprint, INRIA, 1997.
- [Ber96] P. L. Bernstein. *Against the Gods: The Remarkable Story of Risk*. John Wiley & Sons, Inc., 1996.
- [BG98] M. Broadie and P. Glasserman. Simulation for option pricing and risk management. In C. Alexander, editor, *Risk Management and Analysis*. John Wiley & Sons, 1998.
- [BLS00] D. Bertsimas, G. J. Lauprete, and A. Samarov. Shortfall as a risk measure: Properties, optimization and application. Preprint, MIT, 2000.
- [BN77] O.E. Barndorff-Neilsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceeding of the Royal Society London A*, 1977.
- [Bol86] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31(1):34–105, 1986.
- [BS73] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.
- [BTV96] J. Beirlant, J.L. Teugels, and P. Vynckier. *Practical Analysis of Extreme Values*. Leuven University Press, 1996.

- [BVW96] L. Belkacem, J. Lévy Véhel, and C. Walter. CAPM, risk and portfolio selection in stable markets. Preprint, INRIA, 1996.
- [CPB00] P. Cizeau, M. Potters, and J-P Bouchaud. Correlations of extreme stock returns within a non-Gaussian one-factor model. Preprint, Science & Finance, 2000.
- [DdHR00] H. Drees, L. de Haan, and S. Resnick. How to make a Hill plot. *The Annals of Statistics*, 28(1):254–274, 2000.
- [DdV97] J. Danielsson and C. de Vries. Beyond the sample: Extreme quantile and probability estimation. Preprint, LSE, 1997.
- [DdV00] J. Danielsson and C. de Vries. Value at risk and extreme returns. In P. Embrechts, editor, *Extremes and Integrated Risk Management*, pages 85–106. Risk Books, 2000.
- [DHPV01] J. Danielsson, L. De Haan, L. Peng, and C.G. De Vries. Using a bootstrap method to choose the sample fraction in the tail index estimation. *Journal of Multivariate Analysis*, 76:226–248, 2001.
- [DHV98] J. Danielsson, P. Hartmann, and C.G. De Vries. The cost of conservatism: Extreme returns, value at risk and the Basle multiplication factor. *RISK*, 11:101–103, 1998.
- [Dow98] K. Dowd. *Beyond Value at Risk: The New Science of Risk Management*. John Wiley & Sons, Inc., 1998.
- [DP97] D. Duffie and J. Pan. An overview of value at risk. *Journal of Derivatives*, 4(3):7–49, 1997.
- [DSS00] F.X. Diebold, T. Schuermann, and J.D. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance*, 1(Winter):30–36, 2000.
- [EK95] E. Eberlein and U. Keller. Hyperbolic distributions in finance. *Bernoulli*, 1:281–299, 1995.
- [EKM97] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer Berlin, 1997.
- [ELM01] P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. 2001. In this volume.
- [Emb00] P. Embrechts. Extreme value theory: Potential and limitations as an integrated risk management tool. *Derivatives Use, Trading & Regulation*, 6:449–456, 2000.
- [EMS99] P. Embrechts, A.J. McNeil, and D. Straumann. Correlation: Pitfalls and alternatives. *RISK*, 12(5):69–71, 1999.
- [EMS01] P. Embrechts, A.J. McNeil, and D. Straumann. Correlation and dependence in risk management: Properties and pitfalls. In M. Dempster and H.K. Moffatt, editors, *Risk Management: Value at Risk and Beyond*. Cambridge University Press, 2001. To appear.
- [Eng82] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008, 1982.
- [EP00] E. Eberlein and K. Prause. The generalized hyperbolic model: Financial derivatives and risk measures. Preprint, FDM, University of Freiburg, 2000.
- [ERS99] P. Embrechts, S.I. Resnick, and G. Samorodnitsky. Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3:30–41, 1999.
- [Fam65] E.F. Fama. The behavior of stock market prices. *Journal of Business*, 38(1):34–105, 1965.
- [Fam70] E.F. Fama. Risk, return and equilibrium. *Journal of Political Economy*, 79(1):30–55, 1970.
- [FKN90] K.T. Fang, S. Kotz, and K.W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman & Hall, 1990.

- [Fra79] M. Frank. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, 19:194–226, 1979.
- [FV98] E.W. Frees and E.A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25, 1998.
- [Gou97] C. Gouriéroux. *ARCH Models and Financial Applications*. Springer-Verlag, 1997.
- [Hil75] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174, 1975.
- [HL88] C. Huang and R. Litzenberger. *Foundations of Financial Economics*. North-Holland, N.Y., 1988.
- [HW87] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 2:281–300, 1987.
- [Ing87] J. E. Ingersoll. *Theory of Financial Decision Making*. Rowman & Littlefield, 1987.
- [Joe97] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, 1997.
- [Jor01] P. Jorion. *Value at Risk: The New Benchmark for Controlling Market Risk*. McGraw-Hill, 2 edition, 2001.
- [KG00] E. Këllezi and M. Gilli. Extreme value theory for tail-related risk measures. Preprint, University of Geneva, 2000.
- [Lin65] J. Lintner. The valuation of risky assets and the selection of risky investment in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47:13–37, 1965.
- [Lin00a] F. Lindskog. Linear correlation estimation. Preprint, ETH Zürich, 2000.
- [Lin00b] F. Lindskog. Modelling dependence with copulas. Master’s thesis, ETH Zürich, 2000.
- [LK78] H. Levy and Y. Kroll. Ordering uncertain options with borrowing and lending. *Journal of Finance*, 33:553–573, 1978.
- [LLR83] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, 1983.
- [Man63] B.B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:392–417, 1963.
- [Mar52] H. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.
- [Mar59] H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, Inc., 1959.
- [McN98a] A. McNeil. Calculating quantile risk measures for financial return series using extreme value theory. Preprint, ETH Zürich, 1998.
- [McN98b] A. McNeil. On extremes and crashes. *RISK*, January, 1998.
- [Met96] Risk Metrics. Technical document. Technical report, J.P Morgan, 1996.
- [MF00] A. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7:271–300, 2000.
- [Mil99] T.C. Mills. *The Econometric Modelling of Financial Time Series*. Cambridge University Press, 1999.
- [MPR97] S. Mittnik, M.S. Paoletta, and T. Rachev. Modelling the persistence of conditional volatilities with GARCH-stable processes. Preprint, University of California, Santa Barbara, 1997.

- [MRP98] S. Mittnik, T. Rachev, and M.S. Paoletta. Stable Paretian modelling in finance: Some empirical and theoretical aspects. In Adler *et al*, editor, *A Practical Guide to Heavy Tails*. Birkhäuser, 1998.
- [MS97] A. McNeil and T. Saladin. The peaks over threshold method for estimating high quantiles of loss distributions. In *Proceedings of the 28th International ASTIN Colloquium*, 1997.
- [MS00] T Mikosch and C. Stărică. Change of structure in financial time series, long range dependence and the GARCH model. Preprint, 2000.
- [Nel99] R.B. Nelsen. *An Introduction to Copulas*. Springer, New York, 1999.
- [Nol01] J. Nolan. *Stable Distribution: Models for Heavy-Tailed data*. Birkhäuser, 2001.
- [oBS5a] Basle Committee on Banking Supervision. An internal model-based approach to market risk capital requirements. Technical report, Balse, Switzerland: Basle Committee on Banking Supervision, 1995a.
- [oBS5b] Basle Committee on Banking Supervision. Planned supplement to the capital accord to incorporate market risks. Technical report, Balse, Switzerland: Basle Committee on Banking Supervision, 1995b.
- [PMR95] A.K Panorska, S. Mittnik, and T. Rachev. Stable GARCH models for financial time series. *Applied Mathematics Letters*, 8:33–37, 1995.
- [RM00] S. Rachev and S. Mittnik. *Stable Paretian Models in Finance*. John Wiley & Sons, 2000.
- [Ros76] S.A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976.
- [RS97] S. Resnick and C. Stărică. Smoothing the Hill estimator. *Advances in Applied Probability*, 29, 1997.
- [RT01] R-D Reiss and M. Thomas. *Statistical Analysis of Extreme Values*. Birkhäuser, 2 edition, 2001.
- [Sha64] W.F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442, 1964.
- [Shi99] A.N. Shiryaev. *Essentials of Stochastic Finance*. World Scientific, 1999.
- [Sk196] A. Sklar. Random variables, distribution functions, and copulas - a personal look backward and forward. In Rüschendorf *et al*, editor, *Distributions with Fixed Marginals and Related Topics*. Hayward, CA. Institute of Mathematical Sciences, 1996.
- [Smi00] R. Smith. Measuring risk with extreme value theory. In P. Embrechts, editor, *Extremes and Integrated Risk Management*, pages 19–35. Risk Books, 2000.
- [ST94] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, 1994.
- [Sta97] G. Stahl. Three cheers. *RISK*, 10:67–69, 1997.
- [Wil98] T.C. Wilson. Value at risk. In C. Alexander, editor, *Risk Management and Analysis*. John Wiley & Sons, 1998.

Brendan O. Bradley and Murad S. Taqqu
 Department of Mathematics and Statistics
 Boston University
 111 Cummington Street
 Boston, MA 02215
email: bbradley@bu.edu
email: murad@bu.edu
web: <http://math.bu.edu/people/murad>