Dissertation
an der
Fakultät für Mathematik, Informatik und Statistik
der
Ludwig–Maximilians–Universität München

Risk Management Beyond Correlation

vorgelegt von

Tina Yener

München, den 07. Februar 2011
Rigorosum: 05. Dezember 2011

Erstgutachter: Prof. Stefan Mittnik, PhD
Zweitgutachter: Prof. Dr. Martin Missong

Risk Management Beyond Correlation

Tina Yener

In all affairs it's a healthy thing now and then
to hang a question mark on the things
you have long taken for granted.

Bertrand Russell

Abstract

Effects of dependencies are an important component of risk management in an aggregate setting, for example, when different risk types or business lines are to be modeled jointly. Classical approaches to portfolio theory rely on the notion of linear correlation and thus on the ellipticity of the underlying distribution—a requirement which is typically not fulfilled in a risk management context.
For each of market, credit and operational risks, this thesis examines situations where linear correlation reaches its limits and may lead to an underestimation of or counterintuitive effects on aggregate risk. Copulas and tail dependence coefficients are considered, estimated and analyzed as alternatives to correlation. These measures are able to detect extremal dependence structures otherwise neglected. However, due to restrictions by sample sizes and asymptotic considerations, there is no "one–size–fits–all" approach to measuring dependencies among extremes. Instead, this thesis suggests a joint consideration of parametric and nonparametric methods in order to gain a picture of aggregate risk as comprehensive as possible.

Zusammenfassung

Abhängigkeitseffekte stellen einen wichtigen Baustein eines aggregierten Risikomanagement–Prozesses dar, z. B., wenn unterschiedliche Risikotypen oder Geschäftsbereiche gemeinsam modelliert werden. Die klassischen Ansätze der Portfoliotheorie beruhen auf der Idee der linearen Korrelation und damit auf der Elliptizität der zugrunde liegenden Verteilung—eine Voraussetzung, welche typischerweise im Risikomanagement–Kontext nicht erfüllt ist.
Für Markt–, Kredit– und operationelle Risiken betrachtet die vorliegende Arbeit jeweils Situationen, in welchen die lineare Korrelation an ihre Grenzen stößt und zu einer Unterschätzung von oder kontraintuitiven Effekten auf das Gesamtrisiko führen kann. Copulas und Tail Dependence–Koeffizienten werden als Alternativen zur Korrelation betrachtet, geschätzt und analysiert. Diese Maße sind in der Lage, anderenfalls unberücksichtigte Abhängigkeitsstrukturen in den Extrembereichen zu erfassen. Aufgrund der Restriktionen durch Stichprobengrößen und asymptotische Betrachtungen ergibt sich jedoch kein allgemein bester Ansatz zur Messung von Abhängigkeiten zwischen Extremen. Vielmehr schlägt die vorliegende Arbeit eine gemeinsame Betrachtung parametrischer und nichtparametrischer Methoden vor, um ein möglichst umfassendes Bild des Gesamtrisikos zu erhalten.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Risk measurement is a central component of the risk management process. It aims at quantifying the exposure towards different types of risk, such as market, credit, liquidity or operational risks. Typically, an institution will hold a portfolio of risks, that is, it will be exposed to more than only one risk type. Similarly, considering one risk type, it will typically consist of a portfolio of exposures, such as single stocks, credit counterparties or technical devices. In these cases, dependencies are a crucial element of risk quantification: The risk of a portfolio consists of the individual risks and their relationships. Within one risk type, dependencies are typically driven by common factors, such as market conditions for financial risks or an institution's business policy for operational risks. Similarly, on an enterprise level, the different risk types are linked, for example, by the general economic environment. It is therefore of paramount importance to consider dependencies during the risk quantification process. Over the years, many of these linkages have gained relevance due to increasing globalization as well as a higher complexity of derivative and structured products. In view of this development, Enterprise Risk Management (ERM)[1] has recently gained attention. It constitutes a holistic approach to risk management, taking portfolio effects explicitly into account by modeling the aggregate loss distribution on an enterprise level.

The most popular methods of introducing dependencies are based on linear correlation, leading to elegant, closed–form solutions for quantiles and expected values of the loss distribution. In portfolio theory, the approach of Markowitz (1952), the CAPM of Sharpe (1964), Lintner (1965) and Mossin (1966) and the Arbitrage Pricing Theory (APT, Ross, 1976) are based on correlation. In the risk–management framework, the central credit risk model of Merton (1974) as well as the industry models derived from it assume a Gaussian dependence structure. In fact, this assumption of multivariate normality—or, more generally, ellipticity of the joint distribution—is necessary

[1] For an overview of ERM, see, for example, Enterprise Risk Management Committee (2003).

in order for linear correlation to be a meaningful measure of dependence. However, very often the distributions of returns (or losses), from which risk–capital estimates are derived, are far from elliptic. The consequences of such a misspecification depend on the true underlying multivariate structure. For example, in the presence of tail dependence, the Gaussian assumption underestimates the probability of joint extreme movements and will therefore lead to an unrealistically low risk–capital estimate. This may even endanger the existence of an institution in case of occurrence of such extreme movements.

In an ERM framework, aggregation of risks plays an important role. In general, two different roads can be taken. Firstly, in a "bottom–up" approach, individual risk types—or elements of a portfolio within one risk type—may be modeled separately. In the sequel, the resulting risk–capital estimates need to be aggregated, which is most easily done by summation of capital figures. This strategy implicitly assumes a joint occurrence of the adverse movements on which individual capital estimation is based. A second approach, taking a "bird's eye view" perspective, models individual risks jointly. In contrast to the first method, dependencies are explicitly taken into account when constructing the probability distribution of aggregate losses. Risk–capital estimates are then derived directly from the aggregate distribution and do not need to be further aggregated. A comparison of these different strategies leads to non–trivial questions. Firstly, the role of dependencies in the "bird's eye view" approach is not always clear, in the sense that stronger dependencies will imply a higher risk–capital estimate. Secondly, and contrary to intuition as well, the sum of risk–capital estimates may not be a "worst–case–scenario" in the sense that it constitutes an upper bound to the second approach.

In the classical portfolio theory of Markowitz (1952), diversification occurs if the individual components of a portfolio are less than perfectly positively correlated. In this case, the risk of the portfolio is less than the sum of the individual risks, as they partially offset each other. The sum of individual risks thus constitutes a worst–case–scenario. This is why one intuitively expects the "bottom–up" approach of ERM to be more conservative than an aggregate loss modeling, and its result to provide an upper bound to aggregate risk. However, this notion again depends crucially on the assumption of elliptical distributions. In fact, the risk resulting from a joint modeling may exceed the sum of the single risk estimates. In cases where risks reinforce each other ore there are no clear–cut distinctions between risks, this may not be unrealistic. Both is true, for example, for market and credit risks during financial crises.

This thesis aims at assessing dependency structures as well as their effects on risk–capital for market, credit and operational risks. Linear correlation and Value–at–Risk,

two concepts relying on the ellipticity of the underlying distributions, are critically analyzed with respect to their appropriateness as well as the consequences of their use.

Chapter 2 lays the theoretical ground for subsequent analyses. Different dependency concepts and measures are introduced, emphasizing copulas and tail dependence due to their importance for risk management. The problem of measuring risk for non–elliptical distributions with non–subadditive risk measures—such as Value–at–Risk—as well as general "worst–case" bounds on the latter are stated. This chapter draws largely on the classical textbooks of Joe (1997) and Nelsen (1999) as well as McNeil et al. (2005). It contributes to existing research by suggesting an additional criterion for copula goodness–of–fit testing based on nonparametric estimators of extremal dependence.

In Chapter 3, the well–known effect of *correlation breakdown* or *contagion* (Longin and Solnik, 2001) for market risks is analyzed. While most existing studies consider changing correlations among different countries, we assess exceedance correlations and the ability of copulas to capture their shape among different asset classes such as stocks, gold and commodities. The sample used includes data from the financial crisis and thus gives an account on changes of relationships among different asset classes during the recent turbulent times.

When assessing credit risk, one often faces the need to simulate defaults of highly–rated counterparties. In Chapter 4, it is shown that the "bottom–up" approach to portfolio credit risk modeling may imply counterintuitive effects of correlation: A higher correlation among binary occurrence indicators may imply a decrease in the number of event occurrences. In contrast to the well–known superadditivity problem of Value–at–Risk, this effect has been pointed out recently by Mittnik and Yener (2009) in an operational–risk setting.

Operational risk is, due to its heterogeneity, typically modeled separately for several event–type and business–line combinations. In order to derive an aggregate estimate of risk, the two different approaches to risk aggregation may be applied in this context. As pointed out, for example, by Embrechts et al. (2002), the lack of subadditivity of Value–at–Risk may imply an increase in risk–capital when changing from a simple summation approach to an explicit modeling of dependencies. Due to the scarcity of operational risk data, it remains to be analyzed whether this effect may be relevant in real–world application. Drawing on a database of operational losses, Chapter 5 contributes by showing that in some cases, risk capital may increase by up to 30%. Furthermore, it is shown that this increase is partially due to simulation. The use of bounds on Value–at–Risk dating back to Makarov (1981) is suggested in order to disentangle computational and superadditivity effects. Chapter 6 concludes.

Chapter 2

Dependence

In the risk management framework, one generally expects dependencies to have a negative effect on returns—or, correspondingly, a positive influence on aggregate losses. That is, we intuitively expect losses of a portfolio to be higher if the components are positively related. Contagion effects such as those of financial crises are generally assumed to have an impact on diversification, implying that neglecting dependencies will lead to an underestimation of risk. In order to analyze such effects, we first need to have a clear notion of dependence.

Two random variables X_1, X_2 with marginals F_1 and F_2 are known to be independent if their joint distribution is given by

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) \; . \tag{2.1}$$

In principle, the concept of dependence covers all relationships for which Equation (2.1) does not hold. In risk management applications, one is typically concerned about positive dependence, as it may lead to a clustering of losses and thus largely determines the upper tail of the aggregate loss distribution. A very general idea of positive dependence can be summed up by saying that large (small) values of one random variable X_1 tend to occur jointly with large (small) values of a second random variable X_2. In order to detect such behavior, one can compare the joint distribution of X_1 and X_2 to the distribution resulting under the assumption of independence.

2.1 Concepts of Positive Dependence

In the following, if not stated otherwise, we consider the so–called Fréchet class of distributions, that is, the set of joint distribution functions with the same marginals F_1, F_2 (Fréchet, 1951). The distribution defined in (2.1) belongs to this class.

Furthermore, we largely restrict attention to positive dependence; the reasoning behind this is the focus of risk management on extreme (portfolio) losses which are typically

due to parallel movements of the prices of portfolio components. Still, in the two–dimensional setting, concepts of negative dependence can be obtained from simply reversing positive dependence (Lehmann, 1966; Joag-Dev and Proschan, 1983).

Probability integral and quantile transforms will repeatedly appear when working with dependencies; they form the basis for copula modeling and random variable simulation. We therefore state the following proposition, a proof of which can be found, for example, in Feller (1971).

Proposition 1. *Let X be a random variable with distribution function F and F^{-1} denote its quantile function:*

$$F^{-1}(\alpha) = \inf\{x|F(x) \geq \alpha\} , \qquad \alpha \in (0,1) . \tag{2.2}$$

Then

1. *for any standard–uniformly distributed $U \sim Unif(0,1)$, $F^{-1}(U) \sim F$,*
2. *if F is continuous, the random variable $F(X)$ is standard–uniformly distributed: $F(X) \sim Unif(0,1)$.*

The notion of "dependence" canbe cast into restrictions on joint and marginal distributions functions in various ways. A detailed account on dependence concepts and orderings can be found in Shaked and Shanthikumar (2006).

Positive Quadrant Dependence (PQD)

A natural and intuitive way to capture positive dependence is to compare the bivariate distribution to the one resulting from independence (Equation (2.1)). This leads to the notion of positive quadrant dependence (Lehmann, 1966). For a random vector (X_1, X_2) with joint distribution function F and margins F_1, F_2, PQD is defined as

$$\Pr[X_1 \leq x_1, X_2 \leq x_2] \geq \Pr[X_1 \leq x_1] \Pr[X_2 \leq x_2] \ . \tag{2.3}$$

This is equivalent to

$$\Pr[X_1 > x_1, X_2 > x_2] \geq \Pr[X_1 > x_1] \Pr[X_2 > x_2] \ . \tag{2.4}$$

in the bivariate case.[1] Therefore, X_1 amd X_2 are PQD if the probability that they are both small (or both large) is as least as high as under independence. If we

[1] In general, for dimensions $d > 2$ the relationships corresponding to (2.3) and (2.4) are not equivalent. Therefore, one obtains two distinct concepts—positive upper orthant dependence (PUOD) and positive lower orthant dependence (PLOD)—when generalizing PQD to higher dimensions.

consider a second bivariate random vector (Y_1, Y_2) with distribution function G and identical marginals, (X_1, X_2) is smaller than (Y_1, Y_2) in the PQD order—expressed as $(X_1, X_2) \leq_{\text{PQD}} (Y_1, Y_2)$—if

$$F(x_1, x_2) \leq G(x_1, x_2) \qquad \text{for all } x_1 \text{ and } x_2 \,. \tag{2.5}$$

The covariances may be written as (Höffding, 1940)

$$\begin{aligned} \text{Cov}[X_1, X_2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1)F_2(x_2))\,\mathrm{d}x_1\mathrm{d}x_2 \,, \\ \text{Cov}[Y_1, Y_2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (G(x_1, x_2) - F_1(x_1)F_2(x_2))\,\mathrm{d}x_1\mathrm{d}x_2 \,, \end{aligned} \tag{2.6}$$

and as the variances of the marginals in the Fréchet class are equal ($\text{Var}[X_i] = \text{Var}[Y_i]$, $i = 1, 2$), it follows directly that PQD has implications for linear correlation, ρ:

$$(X_1, X_2) \leq_{\text{PQD}} (Y_1, Y_2) \;\Rightarrow\; \text{Cov}[X_1, X_2] \leq \text{Cov}[Y_1, Y_2] \Rightarrow \rho_{X_1,X_2} \leq \rho_{Y_1,Y_2} \,. \tag{2.7}$$

Kimeldorf and Sampson (1987) name seven conditions on the Fréchet class of distributions which should be fulfilled in order to consider the underlying random variables to be positively dependent. In fact, their first condition is precisely the requirement of positive quadrant dependence.

Supermodular Dependence (SMD)

A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is supermodular if

$$\phi(\boldsymbol{x}) + \phi(\boldsymbol{y}) \leq \phi(\boldsymbol{x} \wedge \boldsymbol{y}) + \phi(\boldsymbol{x} \vee \boldsymbol{y}) \,, \tag{2.8}$$

with componentwise maximum (minimum) operator $\wedge$ ($\vee$). We say that (X_1, X_2) is smaller than (Y_1, Y_2) in the supermodular order ($(X_1, X_2) \leq_{\text{SM}} (Y_1, Y_2)$) if

$$\text{E}[\phi(X_1, X_2)] \leq \text{E}[\phi(Y_1, Y_2)] \qquad \text{for all supermodular } \phi : \mathbb{R}^2 \to \mathbb{R} \,. \tag{2.9}$$

If we restrict attention to the bivariate case, SMD boils down do PQD: For $d = 2$, as considered, for example, by Tchen (1980),

$$(X_1, X_2) \leq_{\text{SMD}} (Y_1, Y_2) \;\Leftrightarrow (X_1, X_2) \leq_{\text{PQD}} (Y_1, Y_2) \,, \tag{2.10}$$

However, PQD does not necessarily follow from SMD in the general multivariate case.

Tail Monotonicity: Left–tail Decreasing (LTD), Right–tail Increasing (RTI)

Dividing Equation (2.3) by $\Pr[X_1 \leq x_1]$, we see that PQD can alternatively be written as

$$\Pr[X_2 \leq x_2 | X_1 \leq x_1] \geq \Pr[X_2 \leq x_2] \ . \tag{2.11}$$

If we express the right hand side as conditional probability,

$$\Pr[X_2 \leq x_2 | X_1 \leq x_1] \geq \Pr[X_2 \leq x_2 | X_1 \leq \infty] \ , \tag{2.12}$$

we can directly see the relationship to a stronger condition: Two random variables X_1, X_2 are left–tail decreasing if

$$\Pr[X_2 \leq x_2 | X_1 \leq x_1] \geq \Pr[X_2 \leq x_2 | X_1 \leq x_1'] \qquad \text{for all } x_1 \leq x_1' \text{ and } x_2 \ . \tag{2.13}$$

In other words, $\Pr[X_2 \leq x_2 | X_1 \leq x_1]$ is non–increasing in x_1. Accordingly, X_1 and X_2 are right–tail increasing if

$$\Pr[X_2 \leq x_2 | X_1 > x_1] \geq \Pr[X_2 \leq x_2 | X_1 > x_1'] \qquad \text{for all } x_1 \leq x_1' \text{ and } x_2 \ , \tag{2.14}$$

which means that $\Pr[X_2 \leq x_2 | X_1 > x_1]$ is non–decreasing in x_1. The dependence concepts LTD and RTI can be traced back to Lehmann (1966).

Positive Regression Dependence (PRD)

Two random variables X_1, X_2 are positive regression dependent (sometimes also called stochastically increasing) if

$$\Pr[X_2 \leq x_2 | X_1 = x_1] \geq \Pr[X_2 \leq x_2 | X_1 = x_1'] \qquad \text{for all } x_1 \leq x_1' \text{ and } x_2 \ . \tag{2.15}$$

This property dates back to Tukey (1958) and can alternatively be expressed by saying that $\Pr[X_2 \leq x_2 | X_1 = x_1]$ is non–increasing in x_1.

Positive Association (PA)

Two random variables X_1, X_2 are associated (Esary et al., 1967) if

$$\mathrm{Cov}[f(X_1, X_2), g(X_1, X_2)] \geq 0 \tag{2.16}$$

for all pairs of functions f and g which are non–decreasing in each argument. This concept is particularly relevant in reliability theory (Barlow and Proschan, 1981), where joint failures of technical devices are the objects of interest.

Comonotonicity and Countermonotonicity

Two random variables are comonotonic if there exists a random variable Z and increasing functions f_1, f_2 such that

$$(X_1, X_2) \stackrel{d}{=} (f_1(Z), f_2(Z)) \; . \tag{2.17}$$

From the probability integral transform, it follows that condition (2.17) can alternatively be written as

$$(X_1, X_2) \stackrel{d}{=} (F_1^{-1}(U), F_2^{-1}(U)) \tag{2.18}$$

with $U \sim \text{Unif}(0, 1)$. Comonotonicity can be considered the strongest concept of positive dependence, as the realization of X_i is entirely determined by X_j, $i, j = 1, 2$, $i \neq j$. The joint distribution of comonotonic random variables is given by

$$F(x_1, x_2) = \min(F_1(x_1), F_2(x_2)) \; . \tag{2.19}$$

If $(X_1, -X_2)$ are comonotonic, we obtain perfect negative dependence, which is known as countermonotonicity. In this case,

$$(X_1, X_2) \stackrel{d}{=} (g_1(Z), g_2(Z)) \; , \tag{2.20}$$

where one of the functions g_i, $i = 1, 2$ is increasing and the other one is decreasing. Alternatively, a countermonotonic random vector is given by

$$(X_1, X_2) \stackrel{d}{=} (F_1^{-1}(U), F_2^{-1}(1 - U)) \; . \tag{2.21}$$

The joint distribution function for two countermonotonic random variables can be written as

$$F(x_1, x_2) = \max(F_1(x_1) + F_2(x_2) - 1, 0) \; . \tag{2.22}$$

While it is no problem to extend the concept of comonotonicity to $d > 2$ dimensions, the same is not true for countermonotonicity.

Tail Dependence (TD)

Tail dependence, in contrast to the notions of dependence presented so far, is an asymptotic property. It was introduced by Sibuya (1960); other early accounts on this concept of dependence include Geffroy (1958/59), Tiago de Oliveira (1962/63) and Mardia (1964). Tail dependence may be present in the upper or lower tail of a distribution (or both). Two random variables are upper–tail dependent if

$$\lim_{t \to 1^-} \Pr\left[X_1 > F^{-1}(t) | X_2 > F^{-1}(t)\right] > 0 \; . \tag{2.23}$$

Correspondingly, they are lower–tail dependent if

$$\lim_{t\to 0^+} \Pr\left[X_1 < F^{-1}(t) | X_2 < F^{-1}(t)\right] > 0 \,. \tag{2.24}$$

Loosely speaking, tail dependence implies a positive probability of one variable being extreme given that another variable is extreme is positive.

Relationships among Dependence Concepts

Several relationships among these different dependence concepts can be established.

Assuming (Y_1, Y_2) in (2.7) to be independently distributed ($G(x_1, x_2) = F_1(x_1)F_2(x_2)$), one obtains the following implications of PQD:

$$PQD \;\Rightarrow\; \mathrm{Cov}[X_1, X_2] > 0 \Rightarrow \rho_{X_1,X_2} > 0 \,, \tag{2.25}$$

which equally hold for SMD in the bivariate case. Furthermore, as LTD implies that Equation (2.11) holds (and RTI can be similarly shown to imply (2.4)), LTD and RTI imply PQD. The reverse does, however, not hold.

Lehmann (1966) shows that the implications

$$PRD \;\Rightarrow\; LTD \;\Rightarrow\; PQD$$

hold. Similarly, Esary et al. (1967) point out that PQD also follows from PRD via association:

$$PRD \;\Rightarrow\; PA \;\Rightarrow\; PQD \,,$$

and Esary and Proschan (1972) add that PRD leads to PA via LTD and/or RTI (Joe, 1997, see also). Finally, it can be shown that comonotonicity implies PQD as well (Embrechts et al., 2002):

$$\text{Comonotonicity} \;\Rightarrow\; PA \;\Rightarrow\; PQD \,.$$

Obviously, PQD is a rather general dependence concept, unifying several different notions of dependency. At the same time, as can be seen from the implications among dependence concepts, it is weaker than many other concepts. We will use the concept of PQD in the assessment of bounds on the Value–at–Risk in Chapter 5; it may be assumed to be caused any of the other concepts implying PQD, if deemed more appropriate.

2.2 Dependence Measures

The preceding section contained different concepts of positive dependence. The natural next step is to specify a measure of dependency, capturing the amount of dependence in terms of a real number. In this section, the result will always be a scalar value (also called index); in other words, the dependence structure of a given bivariate distribution will be boiled down to one single value. It is intuitively clear that such an approach may miss important features of a dependence *structure* and is, naturally, less flexible and informative than a function describing the dependence structure. The latter approach, leading to the concept of copulas, will be the topic of Section 2.3 due to its importance. There, the relationships between copulas and the "simpler" measures presented here will be explored.

Rényi (1959) lists seven properties that a measure of dependence should ideally have. This collection was modified and amended by Lancaster (1982) and Schweizer and Wolff (1981). Our main goal is to point out the deficiencies of linear correlation, as well as the advantages of alternative measures. We therefore follow Embrechts et al. (2002) in keeping the list (see Table 2.1) as short as possible and restricting attention to four central requirements.

(1) Symmetry: $\kappa(X_1, X_2) = \kappa(X_2, X_1)$

(2) Normalization: $-1 \leq \kappa(X_1, X_2) \leq 1$

(3) Comonotonicity $\Leftrightarrow \kappa(X_1, X_2) = 1$
Countermonotonicity $\Leftrightarrow \kappa(X_1, X_2) = -1$

(4) For a strictly monotonic transformation $T : \mathbb{R} \to \mathbb{R}$,

$$\kappa(T(X_1), X_2) = \begin{cases} \kappa(X_1, X_2) & \text{for } T \text{ increasing,} \\ -\kappa(X_1, X_2) & \text{for } T \text{ decreasing.} \end{cases}$$

Table 2.1: Desirable Properties of a Dependence Measure κ.

Property (1) implies that no direction of causality is modeled. The second property, setting up upper and lower bounds for $\kappa(\cdot, \cdot)$, is sometimes modified. For example, Lancaster (1982) requires that $0 \leq \kappa(X_1, X_2) \leq 1$, thus neglecting negative relationships. Condition (3) requires that the upper (lower) bound is reached only in the case of comonotonicity (countermonotonicity)—the dependence concept which may be considered the strongest one, as it implies that one variable is a deterministic function of

the other. Property (4) implies that the value assigned to a given dependence structure is not changed by a strictly monotonic transformation of a marginal distribution.

2.2.1 Linear Correlation

Linear (or Pearson) correlation is given by

$$\mathrm{Corr}[X_1, X_2] = \rho_{X_1,X_2} = \frac{\mathrm{Cov}[X_1, X_2]}{\sigma_{X_1}\sigma_{X_2}} \,. \tag{2.26}$$

Equation 2.26 directly shows several of the potential problems which may arise from using correlation. Obviously, it is not defined if the variance $\sigma^2_{X_i}$, $i = 1, 2$, tends to infinity. This is a relevant restriction for financial data, as we may observe very heavy–tailed return distributions. Furthermore, linear correlation does not only depend on the covariance, but also on the marginal distributions, and may thus be affected by changes in the latter. Figure 2.1 shows a simple example of a bivariate distribution with $\mathrm{Cov}[X_1, X_2] = 0.5$ and $\sigma_{X_1} = 1$.

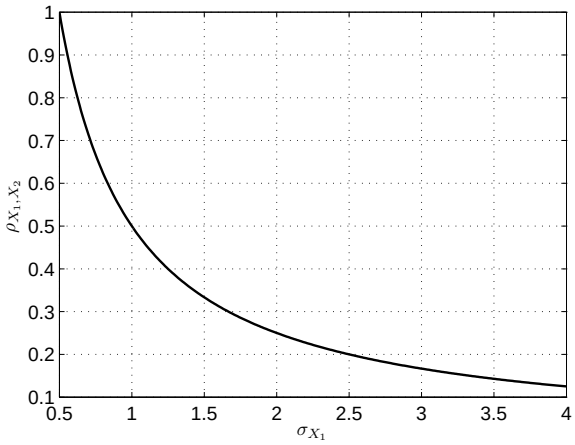


Figure 2.1: Linear Correlation with Changing Standard Deviation

This example illustrates that changes in correlation may be caused by changes in the marginal distributions. This requires caution with respect to interpretations of changes in correlation. For example, in the market risk framework, it is a common perception that correlations increase in times of crises. However, this may, according to Equation 2.26, be due to two reasons: The dependence may have increased, or the variability of returns may have increased (or both). In general, linear correlation is not invariant under nonlinear transformations of the margins.

Further problems arising from the concept of correlation can be summed up in the following theorem.

Theorem 1. *(Höffding, 1940; Fréchet, 1951) Let X_1, X_2 be random variables with marginals F_1, F_2 and unspecified dependence structure. Assume that $0 < \sigma_{X_1}, \sigma_{X_2} < \infty$. Then*

1. *the set of all possible correlations is a closed interval $[\rho_{min}, \rho_{max}]$, and $\rho_{min} < 0 < \rho_{max}$ holds;*

2. *the correlation $\rho = \rho_{min}$ is attained if and only is X_1 and X_2 are countermonotonic, and $\rho = \rho_{max}$ if and only is X_1 and X_2 are comonotonic;*

3. *$\rho_{min} = -1$ if and only is X_1 and $-X_2$ are of the same type; $\rho_{max} = 1$ if and only is X_1 and X_2 are of the same type.*[2]

The second part of this theorem—the proof of which can be found, for example, in McNeil et al. (2005)—may again be seen intuitively by writing the correlation between X_1 and X_2 as

$$\mathrm{Cov}[X_1, X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F(x_1, x_2) - F_1(x_1)F_2(x_2)\} \, \mathrm{d}x\mathrm{d}y \, . \tag{2.27}$$

For given marginals, the maximum correlation corresponds to the maximum covariance; which is, according to Equation (2.27), attained if $F(x_1, x_2) - F_1(x_1)F_2(x_2)$ is maximal. This is equivalent to saying that X_1 and X_2 need to be comonotonic, that is, $F(x_1, x_2) = \min(F_1(x_1), F_2(x_2))$. Correspondingly, the minimum correlation is obtained if $F(x_1, x_2) - F_1(x_1)F_2(x_2)$ is minimal, which is fulfilled if $F(x_1, x_2) = \max(F_1(x_1) + F_2(x_2) - 1, 0)$—in other words, countermonotonicity—holds. The third part of the theorem implies that a comonotonic (countermonotonic) relationship between random variables which are not of the same type (that is, their relationship is nonlinear) is characterized by a correlation $\rho_{\max} < 1$ ($\rho_{\min} > -1$). One example can be constructed from two lognormal distributions: Assuming that $X_1 \sim \mathrm{LN}(0, 1)$ and $X_2 \sim \mathrm{LN}(0, \sigma_{X_2}^2)$, a closed interval $[\rho_{\min}, \rho_{\max}]$ can be constructed such that both bounds converge towards zero as σ_{X_2} increases; see, for example, Embrechts et al. (2002). Another example is given by the Farlie–Gumbel–Morgenstern distribution

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) \left[1 + \alpha(1 - F_1(x_1))(1 - F_2(x_2))\right] \tag{2.28}$$

with $\alpha \in [-1, 1]$ (Farlie, 1960; Gumbel, 1960; Morgenstern, 1956). As shown by Schucany et al. (1978), this distribution implies a correlation coefficient within the range $(-1/3, 1/3)$.

[2] Two random variables V and W are of the same type if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that $V \overset{d}{=} aW + b$.

We can therefore conclude that requirements (3) and (4) of the list in Table 2.1 are not satisfied by linear correlation.

Such examples can be carried forward to show that there may exist situations characterized by perfect dependence and low (or even zero) correlation. Figure 2.2 shows two comonotonic relationships: In the left part, $X_2 = \exp(X_1)$, and in the right part, $X_2 = X_1^2$, with $X_1 \sim \mathrm{N}(0,1)$. Based on 100,000 bivariate observations, estimation of the linear correlation coefficient leads to $\widehat{\rho} = 0.7569$ and $\widehat{\rho} = 0.0098$.

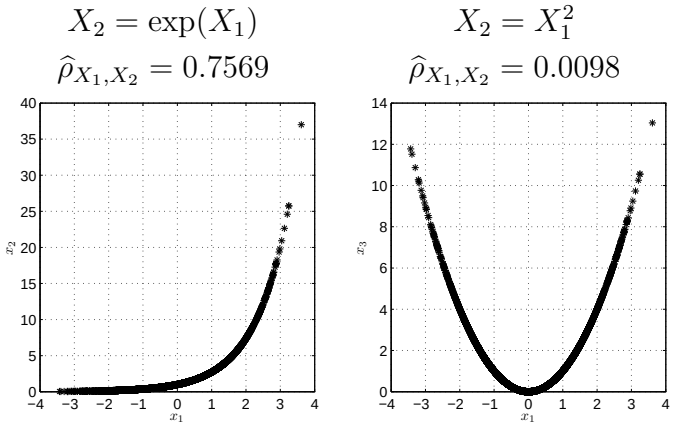
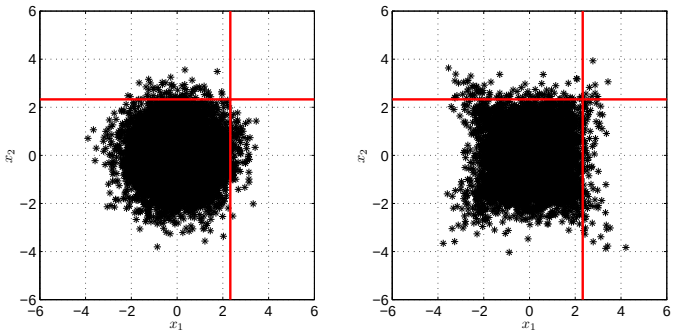


Figure 2.2: Comonotonicity and Linear Correlation

The (theoretical) correlation for $X_2 = X_1^2$ in this setup is equal to zero. This is easily seen by noting that the covariance is given by $\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1^3] - \mathrm{E}[X_1]\,\mathrm{E}[X_2]$, and that the third moment for the standard normal distribution is equal to zero.

These cases illustrate the common and probably most important warning with respect to misinterpretations of correlation: While independence implies a correlation of zero, the reverse does not hold. There are concepts of dependence which correlation cannot capture, as was seen in Section 2.1. This weakness of correlation may be particularly dangerous in a risk management framework, where extreme observations are the focus of analyses. This is illustrated by Figure 2.3, showing two bivariate distributions with a correlation of $\rho_{X_1,X_2} = 0$. Although this value is identical, the joint behavior is obviously different: The scatter plot to the right implies more joint "extreme" observations, that is, observations beyond the 99% quantile of a standard normal random variable for both x_1 and x_2. The reason for this clustering of extreme is a nonzero upper tail dependence in the right plot ($\lambda_U = 0.1326$) implied by a Student–t copula (see Section 2.3) and zero upper tail dependence in the distribution plotted to the left.



Left: 10,000 draws from multivariate normal distribution with $\rho = 0$; right: 10,000 draws from Student–t copula with $\rho = 0$ and $\nu = 5$, using standard normal margins. Red line refers to $\Phi^{-1}(0.99)$.

Figure 2.3: Correlation vs. Tail Dependence

Those joint extreme observations are the ones which one would like to take into account in a model of losses, for example, of a portfolio, as they will largely drive the tail of the distribution of aggregate losses. Neglecting a clustering as in Figure 2.3 by restricting attention to linear correlation may therefore lead to a dramatic underestimation of potential losses.

Correlation and Elliptical Distributions

Correlation is the natural measure of dependency for elliptical distributions, of which the multivariate normal distribution is the most popular representative. However, other elliptical distributions have been applied in finance and risk management; see, for example, Blattberg and Gonedes (1974) and Rachev and Mittnik (2000) for the Student–t and α–stable distributions.

Elliptical distributions are based on the notion of spherical distributions (Fang et al., 1990). A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ has a spherical distribution if

$$U\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X} \tag{2.29}$$

for every orthogonal map $U \in \mathbb{R}^{d\times d}$. This implies that spherical distributions do not change under rotation, as is shown in the left part of Figure 2.4. It contains contour plots of the multivariate normal distribution $\mathrm{N}_2(\boldsymbol{0}, \boldsymbol{I})$. Spherical distributions can be seen as a generalization of the multivariate normal and only represent uncorrelated random variables. However, aside from the multivariate normal distribution, this does not imply independence. Spherical distributions can alternatively defined in terms of

their characteristic function, which is given by

$$\mathrm{E}\left[\mathrm{e}^{i\boldsymbol{t}'\boldsymbol{X}}\right] = \phi(\boldsymbol{t}'\boldsymbol{t}) = \phi(t_1^2 + \cdots + t_d^2) \tag{2.30}$$

and called the characteristic generator. With spherical distribution of $\boldsymbol{X}$ (written as $\boldsymbol{X} \sim \mathrm{S}_d(\phi)$), $\boldsymbol{Y}$ has an elliptical distribution (written as $\boldsymbol{Y} \sim \mathrm{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$) if

$$\boldsymbol{Y} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{X}, \tag{2.31}$$

where $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$. Due to this affine transformation, the contours are given by ellipsoids as shown in the right part of Figure 2.4.[3] The characteristic function of $\boldsymbol{Y}$ is now given by

$$\begin{aligned} \mathrm{E}\left[\mathrm{e}^{i\boldsymbol{t}'\boldsymbol{Y}}\right] &= \mathrm{E}[\exp(i\boldsymbol{t}'(\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{X}))] \\ &= \exp(it'\boldsymbol{\mu})\mathrm{E}[\exp(i(\boldsymbol{A}'\boldsymbol{t})'\boldsymbol{X})] \\ &= \exp(it'\boldsymbol{\mu})\phi(\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}) \ , \end{aligned} \tag{2.32}$$

where the last line follows due to the spherical distribution of $\boldsymbol{X}$ and $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$. The characteristic function of $\boldsymbol{Y} - \boldsymbol{\mu}$ is thus given by $\phi(\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})$. In the case where the second moments are finite, it is always possible to find a representation such that $\boldsymbol{\Sigma} = \mathrm{Cov}[\boldsymbol{Y}]$.

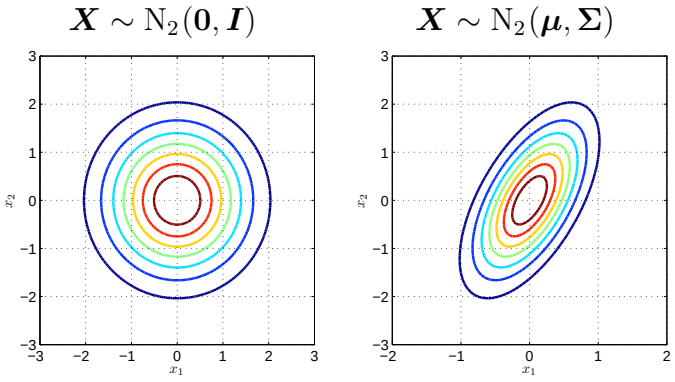


Figure 2.4: Spherical and Elliptical Distributions

From (2.31), it can be directly seen that the multivariate normal distribution $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ belong to the class of elliptical distributions. As is the case with the multivariate normal, linear combinations of elliptical random variables are also elliptical with identical generator ϕ. For $\boldsymbol{Y} \sim \mathrm{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, Kendall's τ (see Section 2.2.2) is given by

$$\tau(X_1, X_2) = \frac{2}{\pi}\arcsin\left(\rho_{X_i,X_j}\right) \ , \tag{2.33}$$

[3] In the one–dimensional case $d = 1$, the class of elliptical distributions coincides with the class of symmetric distributions.

which allows estimation of linear correlation via the more robust rank correlation estimate (Lindskog, 2000).

In the linear world of elliptical distributions, correlation is a natural measure of dependency.[4] This is due to the fact that elliptical distributions are uniquely determined by their means, covariances, and characteristic generators.

2.2.2 Concordance Measures

Intuitively, concordance refers to a notion of dependence among random variables which implies that large values of one random variable tend to occur jointly with large values of the other random variable. That is, a pair of random variables $(x_1, x_2)_i$ and $(x_1, x_2)_j$ are concordant if $x_{1,i} < x_{1,j}$ and $x_{2,i} < x_{2,j}$; correspondingly, discordance is characterized by $x_{1,i} < x_{1,j}$ and $x_{2,i} > x_{2,j}$ or $x_{1,i} > x_{1,j}$ and $x_{2,i} < x_{2,j}$. This implies that two random variables are concordant if $(x_{1,i} - x_{1,j})(x_{2,i} - x_{2,j}) > 0$ and discordant if $(x_{1,i} - x_{1,j})(x_{2,i} - x_{2,j}) < 0$.

Considering a random vector $\boldsymbol{X} = (X_1, X_2)'$ and an independent copy $\boldsymbol{X}' = (X_1', X_2')$, Kendall's τ is defined as

$$\begin{aligned} \tau(X_1, X_2) &= \Pr[(X_1 - X_1')(X_2 - X_2') > 0] - \Pr[(X_1 - X_1')(X_2 - X_2') < 0] \\ &= \mathrm{E}[\,\mathrm{sign}((X_1 - X_1')(X_2 - X_2'))]\ , \end{aligned} \tag{2.34}$$

where the second line follows directly from the definition of the expected value of a discrete random variable. With another independent copy $\boldsymbol{X}'' = (X_1'', X_2'')$, Spearman's ρ_S is given by

$$\rho_S = 3\left(\Pr[(X_1 - X_1')(X_2 - X_2'') > 0] - \Pr[(X_1 - X_1')(X_2 - X_2'') < 0]\right)\ . \tag{2.35}$$

The nature of rank correlations can be illustrated considering a bivariate distribution $F(x_1, x_2)$ with marginals $F_1(x_1)$ and $F_2(x_2)$. Assume that we divide the range into four quadrants using thresholds t_1 and t_2, as illustrated in Figure 2.5. The probabilities of the quadrants are given by

$$\begin{aligned} p_{11} = F(t_1, t_2)\,, \qquad p_{12} = F_1(t_1) - p_{11}\ , \\ p_{21} = F_2(t_2) - p_{11}\,, \qquad p_{22} = 1 - p_{11} - p_{12} - p_{21}\ . \end{aligned}$$

Denoting by p_{i+} a probability summed over random variable j, $j = 1, 2$, we obtain

$$\tau = \frac{p_{22} - p_{2+}p_{+2}}{\sqrt{p_{2+}(1 - p_{2+})}\sqrt{p_{+2}(1 - p_{+2})}}\ , \tag{2.36}$$

[4] For other dependence measures in elliptical distributions, see, for example, Hult and Lindskog (2002).

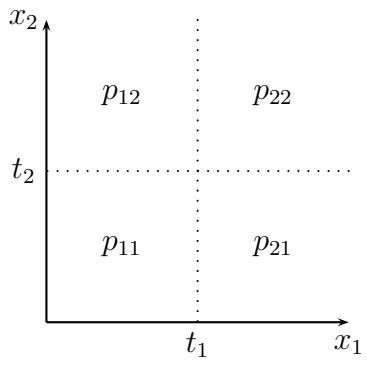
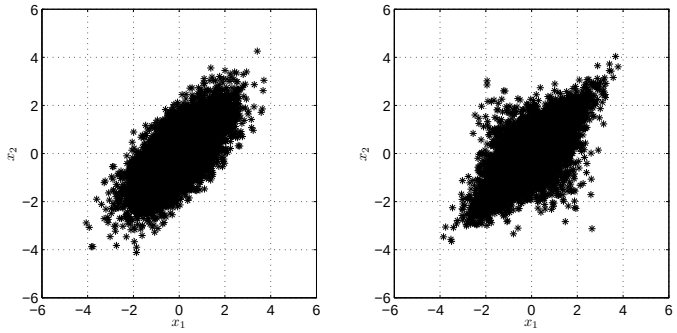


Figure 2.5: Illustration: Concordance and Discordance

which shows that τ is the correlation of ranks (expressed as probabilities).

In contrast to linear correlation, τ and ρ_S are therefore invariant under strictly increasing transformations of the margins. Furthermore, $\tau = 1$ and $\rho_S = 1$ ($\tau = -1$ and $\rho_S = -1$) if random variables X_1 and X_2 are comonotonic (countermonotonic), so that rank correlations fulfill all properties listed in Table 2.1.



Left: 10,000 draws from multivariate normal distribution with $\rho = 0.7071$; right: 10,000 draws from Student–t copula with $\rho = 0.7071$ and $\nu = 5$, using standard normal margins.

Figure 2.6: Rank correlation vs. tail dependence.

However, as rank correlations are scalar values, they are—just as linear correlations—unable to convey information about the *structure* of the joint distribution. For example, the bivariate distributions shown in Figure 2.6 have the same value of Kendall's τ ($\tau = 0.5$) although the behavior in the tails is obviously different due to tail (in)dependence.

For other measures of association, such as Gini's γ and Blomqvist's β, see Nelsen (1999).

2.2.3 Tail Dependence

Tail dependence, as introduced in Equations (2.23) and (2.24), refers to probabilities of joint exceedances of extremes. The natural measure for this asymptotic type of dependence is given by the limits of such exceedance probabilities. Defining the generalized inverse of a distribution function F as

$$F^{-1}(u) = \inf\{x \in \mathbb{R} | F(x) \geq u\} \ , \tag{2.37}$$

the lower tail dependence coefficient is given by

$$\lambda_L = \lim_{t \to 0^+} \Pr[F_1(X_1) \leq t | F_2(X_2) \leq t] \ . \tag{2.38}$$

Correspondingly, the upper tail dependence coefficient is defined as

$$\lambda_U = \lim_{t \to 1^-} \Pr[F_1(X_1) > t | F_2(X_2) > t] \ . \tag{2.39}$$

Tail dependence coefficients are measures of joint extremes and thus concentrate on the distributional area most important to risk management. Typically, either upper or lower tail dependence is in the focus, depending on whether returns or loss distributions are analyzed. Upper and lower tail dependence coefficients may differ, that is, the behavior of joint positive and extremes may be asymmetric.

As tail dependence and its implications are central to the empirical analyses and their interpretations, it will be discussed in more detail in the context of copulas.

2.3 Copulas

Copulas can be viewed as an alternative to the dependence measures presented so far. Instead of boiling down dependency into one single number, copulas are functions which contain the *structure* of a joint distribution. Detailed accounts on the theory of copulas can be found in Nelsen (1999) and Joe (1997).

Formally, a copula can be defined as follows.

Definition 1. *A copula is a function C of d variables on the unit d–cube $[0,1]^d$ with the following properties.*

1. *The range of C is the unit interval $[0,1]$.*
2. *$C(\boldsymbol{u})$ is equal to zero for all $\boldsymbol{u}$ in $[0,1]^d$ for which at least one coordinate is equal to zero.*
3. *$C(\boldsymbol{u}) = u_j$ if all coordinates of $\boldsymbol{u}$ are equal to one except u_j.*

4. *C is d–increasing in the sense that for every $\boldsymbol{a} \leq \boldsymbol{b}$ in $[0,1]^d$, the volume assigned by C to the d–box $[\boldsymbol{a}, \boldsymbol{b}] = [a_1] \times \cdots [a_d, b_d]$ is nonnegative.*

The probably most important theorem in the theory of copulas is Sklar's Theorem. In the following, $\bar{\mathbb{R}}^n$ denotes the extended real line $[-\infty, \infty]$, and Ran F refers to the range of the distribution function F.

Theorem 2. *(Sklar, 1959) Let F be a d–dimensional distribution function with margins $F_1, \ldots, F_n$. Then there exists a copula C such that for all $\boldsymbol{x}$ in $\bar{\mathbb{R}}^d$,*

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_n(x_d)\right) \ . \tag{2.40}$$

If $F_1, \ldots, F_d$ are all continuous, C is unique; otherwise it is uniquely determined on Ran $F_1 \times \cdots \times$ Ran F_d. Conversely, if C is a copula and $F_1, \ldots, F_d$ are distribution functions, then the function F is a d–dimensional distribution function with margins $F_1, \ldots, F_d$.

Summing up, a copula is a multivariate distribution function with univariate margins being uniformly distributed between zero and one. It couples univariate distribution functions (which are known to be uniformly distributed according to the probability–integral transform), and its outcome forms the joint distribution function of the underlying variables. In general, copulas can be defined for dimensions $d > 2$; however, we restrict attention to the bivariate case. Furthermore, we assume the distribution functions of the random variables under consideration to be continuous.[5] Sklar's Theorem, which illustrates how a copula "couples" univariate margins to their joint distribution function, can alternatively be written as

$$C = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right) \ , \tag{2.41}$$

where $U_1, U_2 \sim \text{Unif}(0,1)$ and the inverses F_1^{-1}, F_2^{-1} are assumed to exist.

An attractive property of copulas consists of their invariance towards strictly increasing transformations of the marginals $F_1(x_1)$, $F_2(x_2)$. This is due to the fact that, just like rank correlations, copulas are based on the ranks of random variables. In fact, rank correlations only depend on the copula of the underlying data. For Kendall's τ, this can be shown taking into account that for continuous random vectors $\boldsymbol{X}$ and $\boldsymbol{X}'$,

$$\Pr[(X_1 - X_1')(X_2 - X_2') > 0] + \Pr[(X_1 - X_1')(X_2 - X_2') < 0] = 1 \ . \tag{2.42}$$

[5] In principle, copulas can still be worked with in the discrete case; however, non–continuous margins pose special challenges. This topic is treated, for example, by Marshall (1996) and Genest and Nešlehová (2007).

Equation (2.34) can thus be rewritten as

$$\begin{aligned}\tau(X_1, X_2) &= 2\Pr[(X_1 - X_1')(X_2 - X_2') > 0] - 1 \\ &= 4\Pr[X_1 < X_1', X_2 < X_2'] - 1 \\ &= 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x_1, x_2)dF(x_1, x_2) - 1 \\ &= 4\int_0^1\int_0^1 C(u_1, u_2)\mathrm{d}C(u_1, u_2) - 1 \,,\end{aligned}$$

where the second line follows from the fact that the pairs are independent. This result implies that a value for τ automatically follows from a given copula $C(\cdot, \cdot)$.

Any copula is bounded from below and above. These lower and upper bounds, $C_\ell(u_1, u_2)$ and $C_u(u_1, u_2)$, correspond to the cases of countermonotonicity and comonotonicity. They date back to Fréchet (1951) and Höffding (1940) and are therefore known as Frechet–Höffding bounds:

$$\underbrace{\max(u_1 + u_2 - 1, 0)}_{C_\ell(u_1,u_2)} \leq C(u_1, u_2) \leq \underbrace{\min(u_1, u_2)}_{C_u(u_1,u_2)} \,. \tag{2.43}$$

Another so–called fundamental copula is given by the independence copula

$$C_I(u_1, u_2) = u_1 u_2 \,, \tag{2.44}$$

whose contour plots and surfaces are shown in Figure 2.7 jointly with those of C_ℓ and C_u.

Copulas may equally be applied to survival functions $\bar{F}(x) = \Pr[X > x] = 1 - F(x)$. Using the joint survival function

$$\bar{F}(x_1, x_2) = \Pr[X_1 > x_1, X_2 > x_2] = 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2) \,, \tag{2.45}$$

one obtains a relationship as in Equation (2.40):

$$\bar{F}(x_1, x_2) = \bar{C}\left(\bar{F}_1(x_1), \bar{F}_2(x_2)\right) \,, \tag{2.46}$$

where

$$\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2) \tag{2.47}$$

is the survival copula of X and Y.[6]

[6]The survival copula must not be confused with the joint survival probability

$$C^*(u_1, u_2) = \Pr[U_1 > u_1, U_2 > u_2] = 1 - u_1 - u_2 + C(u_1, u_2) \,,$$

corresponding to $\bar{F}(\cdot, \cdot)$ from Equation (2.45).

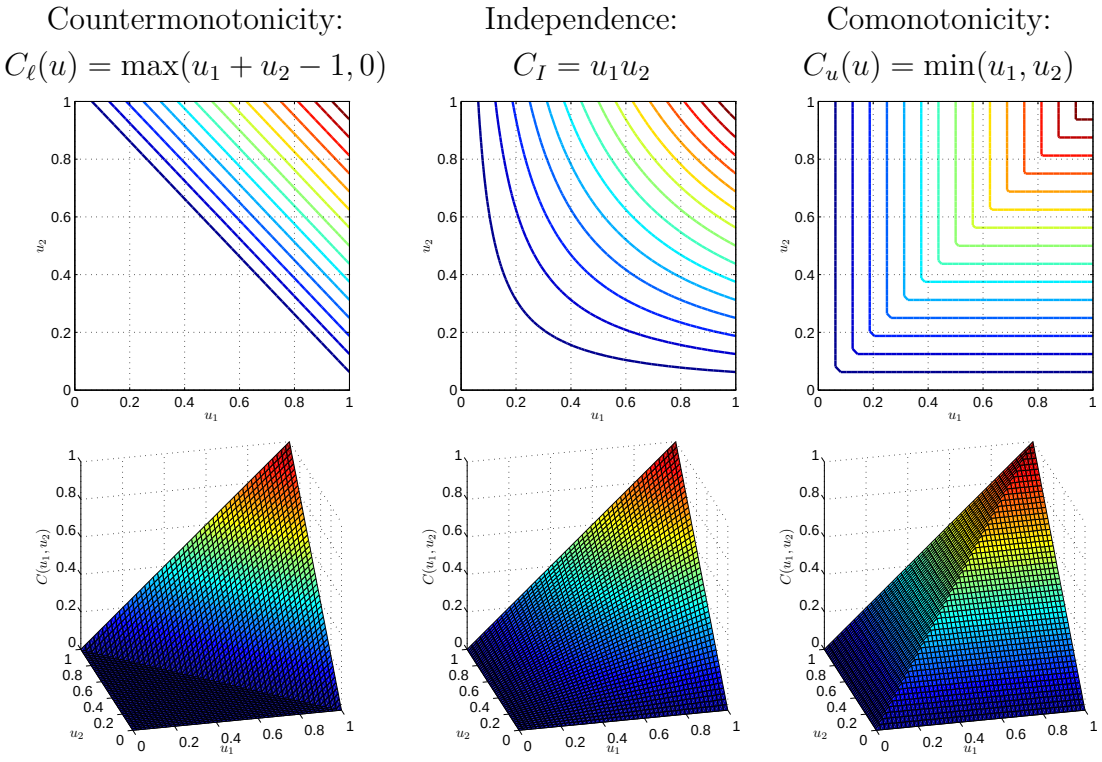


Figure 2.7: Fundamental Copulas

The density of a copula is given by

$$c(u_1, u_2) = \frac{\partial C(u_1, u_2)}{\partial u_1 \partial u_2} \,. \tag{2.48}$$

For survival copulas, the relationship

$$\tilde{c}(u_1, u_2) = c(1 - u_1, 1 - u_2) \tag{2.49}$$

illustrates the fact that they are mirror images of $C(u_1, u_2)$. An illustration of this will be given below.

Elliptical Copulas

Elliptical copulas are copulas of elliptical distributions. The two most famous members of this class are the Gaussian and the Student–t copula. The former is given by

$$\begin{aligned} C^{\mathrm{Ga}}(u_1, u_2) &= \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{2\rho z_1 z_2 - z_1^2 - z_2^2}{2(1-\rho^2)}\right) \mathrm{d}z_1 \mathrm{d}z_2 \,, \end{aligned} \tag{2.50}$$

where Φ_ρ denotes the cumulative distribution function of the bivariate normal distribution with correlation ρ. As shown by Roncalli (2002), this can be alternatively

expressed as

$$C^{\mathrm{Ga}}(u_1, u_2) = \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}}\right) \mathrm{d}v \ . \tag{2.51}$$

If two variables X_1, X_2 are each normally distributed, and their dependence structure is given by a Gaussian copula with correlation ρ, then their multivariate distribution is normal with correlation ρ. However, if the marginals are not normally distributed, then the resulting bivariate distribution is not a multivariate normal. For illustration purposes, Figure 2.8 compares the joint distributions resulting from the same Gaussian copula with linear correlation, but different (standard normal vs. Student–t) margins. Obviously, the resulting distributions are very different, the one on the right being neither multivariate normal nor multivariate t. Its margins are heavy–tailed, but its joint distribution does not exhibit any tail dependence. This demonstrates the copula's ability of splitting a multivariate distribution into the marginal components and the dependence structure.

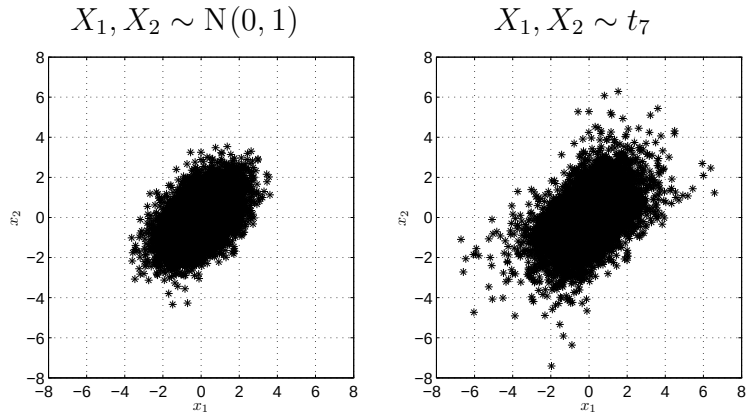


Figure 2.8: Gaussian Copula ($\rho = 0.5$), Normal vs. Student–t margins

Accordingly, the Student–t copula

$$C^t = t_{\rho,\nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) \tag{2.52}$$

is based on $t_{\rho,\nu}$, the bivariate cdf of the Student–t distribution with ν degrees of freedom and correlation ρ. Here, t_ν^{-1} denotes the inverse of the (univariate) Student–t distribution with ν degrees of freedom. As ν increases, the tail dependence of the bivariate distribution decreases, and the copula converges to the Gaussian copula.

As there are no closed–form expressions for these two copulas, the Gaussian and Student-t copulas are sometimes also called implicit copulas.

Archimedean Copulas

Archimedean copulas date back to Schweizer and Sklar (1961) and Ling (1965).[7] They are particularly popular due to their tractability and the large number of different families of copulas contained in this class. Additionally, they are able to model asymmetric tail dependence structures. All Archimedean copulas can be written as

$$C(u_1, u_2) = \phi^{[-1]}(\phi(u_1), \phi(u_2)) \ , \tag{2.53}$$

where $\phi(t) : I \to \bar{\mathbb{R}}^+$ is the continuous, decreasing and convex generator of the copula, satisfying $\phi(1) = 0$ and with pseudo–inverse

$$\phi^{[-1]}(u) = \begin{cases} \phi^{-1}(u) & \text{for} \quad 0 \leq u \leq \phi(0) \ , \\ 0 & \text{for} \quad \phi(0) \leq u \leq +\infty \ . \end{cases} \tag{2.54}$$

If $\phi(0) = +\infty$, the generator is additionally called strict, and its pseudo–inverse corresponds to the usual inverse.[8]

Dependence, as measured by rank correlation or tail dependence, can be directly related to the copula's generator. As shown by Genest and MacKay (1986), Kendall's τ for Archimedean copulas can be written as

$$\tau = 4 \int_I \frac{\phi(u)}{\phi'(u)} \mathrm{d}u + 1 \ . \tag{2.55}$$

The coefficients of upper and lower tail dependence are given by (Joe, 1997)

$$\lambda_U = 2 - 2 \lim_{s \to 0^+} \frac{\phi'(2s)}{\phi'(2s)} \tag{2.56}$$

and

$$\lambda_L = 2 \lim_{s \to +\infty} \frac{\phi'(2s)}{\phi'(s)} \ . \tag{2.57}$$

In our empirical applications, we focus on the two most commonly used families of Archimedean copulas. The Gumbel copula[9] (Gumbel, 1960)—or, due to Hougaard (1986), Gumbel–Hougaard copula—is given by

$$C^{Gu}(u_1, u_2) = \exp\left(- \left[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha \right]^{1/\alpha} \right) \ , \quad \alpha \in [1, \infty) \ . \tag{2.58}$$

[7] Archimedean copulas found first applications in the actuarial field, see Cook and Johnson (1981), Duchateau and Janssen (2008) and Oakes (1982) as well as Frees and Valdez (1999) for an overview.

[8] If this inverse is completely monotone, the corresponding Archimedean copula represent PQD.

[9] The Gumbel copula is a bivariate extreme value copula; see Section 2.3.2 and, for example, Segers (2004) and Balkema and Embrechts (2007).

For $\alpha = 1$, it reduces to the independence copula $C_I = u_1 u_2$, and for $\alpha \to \infty$, it reaches the upper Fréchet bound. The same is true for the limits of the parameter of the Clayton copula (Clayton, 1978)

$$C^{Cl}(u_1, u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-1/\alpha} , \quad \alpha \in [-1, \infty) \backslash \{0\} . \tag{2.59}$$

Additionally, for the Clayton copula the lower Fréchet bound is reached if $\alpha = -1$.

Figure 2.9 illustrates the copulas used in our analyses. The first and second column contain surface and contour plots of the copulas' densities; the third column juxtaposes 10,000 random draws from the corresponding copula. In all cases, Kendall's τ is taken to be $\tau = 0.6$, and the marginals are assumed to be standard–normally distributed. In contrast to the Gaussian and Student–t copulas' symmetry, the Gumbel and Clayton copulas imply asymmetric dependence structures. The Clayton survival copula, as mirror image of the Clayton, shows upper instead of lower tail dependence and can therefore be used as alternative to the Gumbel in order to model positive upper but absence of lower tail dependence. The Gumbel survival copula, which is not depicted, leads to lower tail dependence, accordingly.

Summing up, Tables 2.2 and 2.3 list Kendall's τ and tail dependence coefficients for the copulas used in our analyses.

Copula	Kendall's τ
Gaussian	$\tau = \frac{2}{\pi} \arcsin(\rho)$
Student–t	$\tau = \frac{2}{\pi} \arcsin(\rho)$
Gumbel (Survival)	$\tau = \frac{\alpha - 1}{\alpha}$
Clayton (Survival)	$\tau = \frac{\alpha}{2 + \alpha}$

Table 2.2: Copulas and Kendall's τ

Copulas have been criticized for pretending to deliver easy solutions to many difficult statistical problems without fulfilling this promise. This general worry is expressed, for example, by Mikosch (2006); for a criticism of the Gaussian copula in the context of the financial crisis see Salmon (2009). However, underestimation of risk, or, more generally, failure to model a given multivariate distribution must not attributed to the concept of copulas *per se*. Copulas are useful abstract instruments to be handled with care. For example. a multivariate distribution with substantial tail dependence will not be adequately represented by a Gaussian copula—but the implied underestimation of risk–capital is due to the incorrect *use* of copulas.

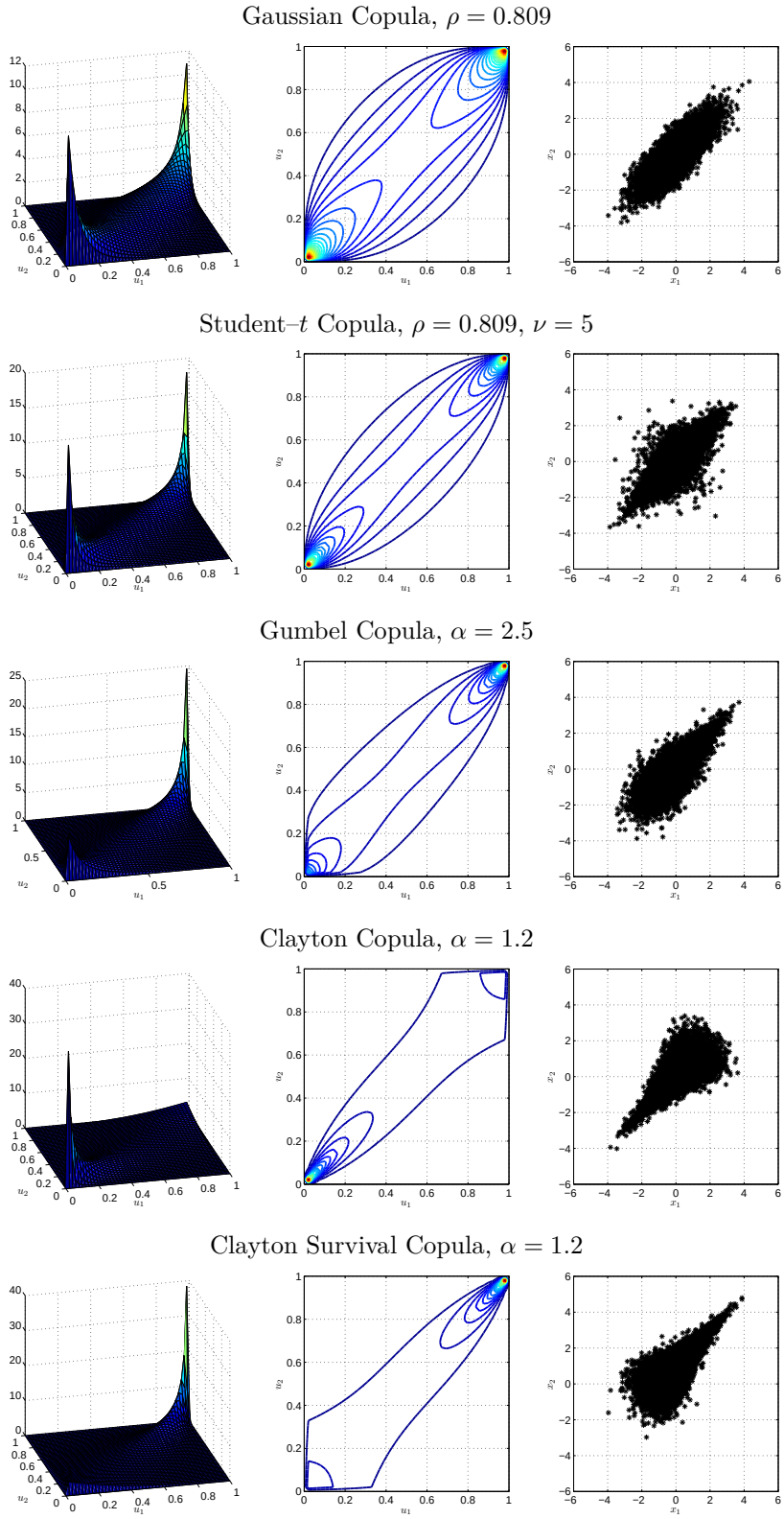


Figure 2.9: Copula Densities and Scatterplots, $\tau = 0.6$

Copula	λ_L	λ_U
Gaussian	0	0
Student–t	$2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}}\right)$	$2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}}\right)$
Gumbel	0	$2-2^{1/\alpha}$
Gumbel Survival	$2-2^{1/\alpha}$	0
Clayton	$2^{-1/\alpha}$	0
Clayton Survival	0	$2^{-1/\alpha}$

Table 2.3: Copulas and Tail Dependence Coefficients

2.3.1 Copula Estimation

Given observations from a multivariate sample, one faces the task of estimating its copula. In principle, this can be done nonparametrically, using the empirical copula of Deheuvels (1981). Denote the sample's order statistics by $\{x_1^{(t)}, x_2^{(t)}\}$ and the corresponding rank statistics by $\{R_1^{(t)}, R_2^{(t)}\}$. The empirical copula is defined—in analogy to the univariate empirical distribution function—as

$$\widehat{C}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) = \frac{1}{T}\sum_{t=1}^{T} \mathcal{I}_{(-\infty,t_1]}\left(R_1^{(t)}\right)\mathcal{I}_{(-\infty,t_2]}\left(R_2^{(t)}\right)\ ,\ t_1 = 0,\ldots,T,\ t_2 = 0,\ldots,T\ , \tag{2.60}$$

where

$$\mathcal{I}_{\mathcal{A}}(x) = \begin{cases} 1 & \text{if} \quad x \in \mathcal{A} \\ 0 & \text{if} \quad x \notin \mathcal{A} \end{cases} \tag{2.61}$$

denotes the indicator function. The empirical copula is thus given by the number of pairs (x_1, x_2) in the sample such that $x_1 \leq x_1^{(t_1)}$ and $x_2 \leq x_2^{(t_2)}$, divided by the total sample size. It can be shown that the empirical copula converges to the true underlying copula (Deheuvels, 1979, 1981).

Although $\widehat{C}\left(\frac{t_1}{T}, \frac{t_2}{T}\right)$ represents the data's dependence structure, estimation of dependence via rank correlations or tail dependence coefficients is, in general, not easily obtained. However, as pointed out by Joe (1997), sample versions of several dependence measures can be obtained from the empirical copula. Just like in the univariate case, the empirical copula can be smoothed via kernel techniques; see, for example, Fermanian and Scaillet (2003).

A second approach towards copula estimation, as suggested by Genest and MacKay (1986), relies on the relationships among copulas and dependence measures established

above. For example, we may estimate Kendall's τ from the sample and infer the corresponding copula parameter using Table 2.2. However, it is not clear which parametric copula to choose, and this is why this approach is only useful when having determined the appropriate copula in a first step. Furthermore, the estimation results depend entirely on a (scalar) coefficient and are obtained neglecting useful dependence structures among the data.

Due to these limitations, we exclusively focus on estimation via Maximum Likelihood.

Maximum Likelihood: IFM

From the copula density in Equation (2.48), it follows that the joint density of X_1 and X_2 can be written as

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2) \ .$$

This shows that inference for the joint distribution involves an identification of the marginal distributions, but also the specification of the copula function to be used. The Log–likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c(F_1(x_{1,t}), F_2(x_{2,t})) + \sum_{t=1}^{T} f_1(x_{1,t}) + \sum_{t=1}^{T} f_2(x_2, t) \ ,$$

where $\boldsymbol{\theta}$ contains the parameters from the univariate distribution and the copula, needs then to be maximized with respect to $\boldsymbol{\theta}$, leading to the exact ML estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$. This can prove to be computationally intensive, especially in dimensions $d > 2$.

Therefore, the so–called *Inference Functions for Margins*, or IFM method, was proposed by Joe and Xu (1996). It divides estimation into two steps: First, the margins' parameters $\boldsymbol{\theta}_{\mathrm{m}} = (\theta_1, \theta_2)$ are estimated as

$$\widehat{\boldsymbol{\theta}}_{\mathrm{m}} = \underset{\boldsymbol{\theta}_{\mathrm{m}}}{\operatorname{argmax}} \left(\sum_{t=1}^{T} f_1(x_{1,t}; \theta_1) + \sum_{t=1}^{T} f_2(x_{2,t}; \theta_2) \right) \ .$$

In a second step, the copula parameter θ_{c} is estimated, given $\boldsymbol{\theta}_{\mathrm{m}}$:

$$\widehat{\theta}_{\mathrm{c}} = \underset{\boldsymbol{\theta}_{\mathrm{c}}}{\operatorname{argmax}} \sum_{t=1}^{T} \ln c(F_1(x_{1,t}), F_2(x_{2,t}); \theta_{\mathrm{c}}, \widehat{\boldsymbol{\theta}}_{\mathrm{m}})$$

The IFM estimator is, accordingly,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{IFM}} = (\widehat{\boldsymbol{\theta}}_{\mathrm{m}}, \widehat{\theta}_{\mathrm{c}}) \ .$$

While for the Gaussian copula with correlation ρ and normally distributed univariate margins $\widehat{\boldsymbol{\theta}}_{\mathrm{IFM}} = \widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$, this equivalence does not hold in general. Nevertheless, Joe

(1997) points out that the IFM method is computationally simpler and highly efficient, compared to exact ML estimation.

Maximum Likelihood: CML

A semiparametric method is the so–called *Canonical Maximum Likelihood*, or CML estimation. The underlying idea is that a misspecification of marginals in the exact ML and IFM methods makes the estimators lose their consistency. In cases where one is purely interested in assessing the dependenc structure, it might thus be preferable not to restrict the marginal behavior. Therefore, the first step of the IFM method is replaced by the calculation of the empirical distribution functions $\hat{F}(x)$ and $\hat{G}(y)$.[10] The copula parameter θ_c is then estimated via Maximum Likelihood and obtained from

$$\widehat{\theta}_{\text{CML}} = \underset{\boldsymbol{\theta}_c}{\text{argmax}} \sum_{t=1}^{T} \ln c(\hat{F}_1(x_{1,t}), \widehat{F}_2(x_{2,t}); \theta_c)$$

Genest et al. (1995) show that under suitable regularity conditions, $\widehat{\theta}_{\text{CML}}$ is consistent, and the quantity $T^{-1/2}(\widehat{\theta}_{\text{CML}} - \theta_c)$ is asymptotically normal. However, this method suffers from a loss of efficiency, see Genest and Rivest (1993).
If the margins are unknown—as is typically the case for empirical analyses—the CML method outperforms exact ML and IFM in terms of mean–squared errors.

This method, which we choose for our empirical applications, represents an appropriate compromise between a fully parametric and a completely nonparametric approach.

2.3.2 Tail Dependence Revisited

We have already seen the relationship between the tail dependence coefficient and the parameters of different copulas in Table 2.3. In fact, the tail dependence coefficient as expressed in Equation (2.39) can alternatively be written in terms of the copula of X_1 and X_2:

$$\lambda_U = \lim_{t \to 1^-} \Pr[F_1(X_1) > t | F_2(X_2) > t] = \lim_{t \to 1^-} \frac{\Pr[F(X) > t, G(Y) > t]}{\Pr[G(Y) > t]} \,. \tag{2.62}$$

Using the probability transform, we write

$$\begin{aligned} \lambda_U &= \lim_{t \to 1^-} \frac{\Pr[U_1 > t, U_2 > t]}{\Pr[U_2 > t]} \\ &= \lim_{t \to 1^-} \frac{1 - F_{U_1}(t) - F_{U_2}(t) + \Pr[U_2 \leq t, U_1 \leq t]}{1 - F_{U_2}(t)} \,, \end{aligned}$$

[10] Genest et al. (1995) recommend to replace the empirical distribution function $\hat{F}$ by $\hat{F}_T$, which is $T/(T-1)$ times $\hat{F}$, in order to avoid problems caused by the marginal cdfs tending to one (as $c(1,1) = 0$). This modification will be used in our empirical analyses.

where F_{U_i}, $i = 1, 2$, denotes the distribution function of U_i. But as this is a uniform distribution, $U_i \sim \text{Unif}(0, 1)$, which implies that $F(u) = u$, this can be written as

$$\lambda_U = \lim_{t \to 1^-} \frac{1 - 2t + C(t, t)}{1 - t} = \lim_{t \to 1^-} \left(2 - \frac{1 - C(t, t)}{1 - t} \right) . \tag{2.63}$$

Defining $s = 1 - t$, this implies that λ_U is the lower tail dependence coefficient of the survival copula $\bar{C}(s, s)$. Similarly, it can be shown that for the lower tail dependence coefficient,

$$\lambda_L = \lim_{t \to 0^+} \frac{C(t, t)}{t} . \tag{2.64}$$

These representations confirm that the tail dependence coefficient is determined by the copula of X and Y and is also invariant under strictly increasing transformations of the margins.

Multivariate Extreme Value Theory

The tail dependence coefficient considers the probability of joint exceedances of high thresholds. This probability is boiled down into one single number and could, therefore, be criticized just as correlations which do not convey much information about the structure of dependence. An improvement can be achieved in the context of multivariate extreme value theory, which additionally implies a relationship between such extremal dependence structures and the tail dependence coefficient.

In univariate extreme value theory, we know that the distribution of exceedances of a high threshold converges to a Generalized Pareto distribution (GPD), implying that the maxima of blocks converge to a Generalized Extreme Value (GEV) distribution (Embrechts et al., 2007). Multivariate extreme value theory (see, for example, Galambos, 1987) aims at finding such domain–of–attraction properties for joint distributions—or copulas—of random variables.

Considering m blocks of length n, their maxima are given by

$$M_{X_1,n} = \max(x_{1,}, \ldots, x_{1,n}) , \qquad M_{X_2,n} = \max(x_{2,}, \ldots, x_{2,n}) , \tag{2.65}$$

and we write $\boldsymbol{M}_n = (M_{X_1,n}, M_{X_2,n})$. The focus of multivariate extreme value theory is on the joint asymptotic behavior of these maxima. That is, one is interested in conditions on the original distribution of $\boldsymbol{X} = (X_1, X_2)$ under which

$$\Pr[\boldsymbol{M}_n \leq \boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n] = F^n(\boldsymbol{a}_n \boldsymbol{x} + \boldsymbol{b}_n) \overset{n \to \infty}{\longrightarrow} G(\boldsymbol{x}) \tag{2.66}$$

holds. In this case, $G(\cdot)$ is a multivariate extreme value distribution (MEVD) and F is in the domain of attraction of G, written as $F \in \mathrm{MDA}(G)$. This setup is identical to the univariate case; however, in the multivariate framework there is no finite–dimensional family covering the entire class of such distributions. In order to keep notation simple and exclude univariate aspects, the marginals are typically standardized using univariate Fréchet distributions.

If a limit distribution G exists, its marginals follow univariate extreme value distributions $G_1(x_1)$ and $G_2(x_2)$. The copula of $G(\boldsymbol{x})$ is thus unique, owing to the continuity of the marginals. From Sklar's theorem, the multivariate extreme value distribution can be written as

$$G(\boldsymbol{x}) = C_e(G_1(x_1), G_2(x_2)) \tag{2.67}$$

where $C_e(\cdot,\cdot)$ denotes an extreme value copula. In this case, the copula C of F is in the domain of attraction of an extreme value copula, $C \in CDA(C_e)$, where

$$C_e(\boldsymbol{u}^t) = C^t(\boldsymbol{u}) \; . \tag{2.68}$$

Any copula for which relationship (2.68) holds is an extreme value copula.[11]

Under some weak regularity conditions, the multivariate extreme value distribution can be written as

$$G(\boldsymbol{x}) = \exp(-V(x_1, x_2)) \tag{2.69}$$

with

$$V(x_1, x_2) = \int_0^1 \max\left(\frac{\omega}{x_1}, \frac{1-\omega}{x_2}\right) 2\mathrm{d}H(\omega) \; , \quad \omega \in [0,1] \; , \tag{2.70}$$

where H is a distribution function on $[0,1]$ satisfying

$$\int_0^1 \omega \mathrm{d}H(\omega) \frac{1}{2} \; . \tag{2.71}$$

Alternatively, one can use the fact that

$$\int_0^1 \max\left(\frac{\omega}{x_1}, \frac{1-\omega}{x_2}\right) H(\mathrm{d}\omega) = \left(\frac{1}{x} + \frac{1}{y} A\left(\frac{x}{x+y}\right)\right) \; , \tag{2.72}$$

where

$$A(\nu) = \int_0^1 \max(\omega(1-\nu), (1-\omega)\nu) H(\mathrm{d}\omega) \tag{2.73}$$

[11] For example, the independence copula as well as the comonotonicity and the Gumbel copula belong to this class.

with

$$\max(\nu, 1-\nu) \leq A(\nu) \leq 1 \tag{2.74}$$

is convex on $[0, 1]$ and known as Pickands' dependence function (Pickands, 1981).

Extreme value copulas can be expressed in terms of Pickand's dependence function: a copula C is an extreme–value copula if and only if it can be written as

$$C(u_1, u_2) = \exp\left\{(\ln(u_1) + \ln(u_2))A\left(\frac{\ln(u_1)}{\ln(u_1) + \ln(u_2)}\right)\right\} . \tag{2.75}$$

By inserting into Equation (2.75), we see that the bounds in equation (2.74) correspond to two fundamental copulas: $A(\nu) = 1$ implies independence ($C(u_1, u_2) = u_1 u_2$), and $A(\nu) = \max(\nu, 1-\nu)$ leads to comonotonicity ($C(u_1, u_2) = \min(u_1, u_2)$).

If $C \in CDA$, the upper tail dependence coefficient of C_e is identical to that of C. Furthermore, it is related to Pickand's dependence function via $\lambda_U = 2(1 - A(1/2))$ (McNeil et al., 2005). This implies that one can—under the assumption of a multivariate extreme distribution—infer Pickand's dependence function and thus the extreme value copula from λ_U.
One important representative of bivariate extreme–value copulas is the Gumbel copula. In fact, Archimedean copulas for which $C \in CDA$ holds are in the domain of attraction of the Gumbel copula. This makes the Gumbel copula a natural candidate for tail dependence estimation when considering Archimedean copulas (see, for example, Capéraà and Fougères, 2000; Genest and Rivest, 1989).

Multivariate extreme value theory has lead to considerations of copulas for the tails related to bivariate extreme value copulas—just as exceedances of random variables can be used in the context of univariate Extreme Value Theory. For example, Juri and Wüthrich (2002) define a so–called upper tail copula, corresponding to the lower threshold copula of McNeil et al. (2005).

Asymptotic versus Quantile Dependence

The tail dependence coefficient from Equation (2.63) is defined as the limit for $t \to 1$, that is, it provides information with regard to asymptotic dependence. Standard multivariate extreme value theory methods as presented above are based on distributions for which either $\lambda_U > 0$ (asymptotic dependence) or independence of the entire distribution (implying $\lambda_U = 0$) holds. Such methods are, thus, not able to model distributions which are asymptotically independent but show extremal dependence for $t < 1$. This is a relevant limitation because empirical analyses often lead to $\lambda_U = 0$, in which case an analysis of dependence at lower quantiles is still desirable. The multivariate normal

distribution with $\rho < 1$ is an example of an asymptotically independent distribution showing such quantile dependence (Galambos, 1987; Embrechts et al., 2002). Methods for this case have been developed by Coles and Tawn (1994), Ledford and Tawn (1996), Ledford and Tawn (1997), Bruun and Tawn (1998) and Bortot and Tawn (1998).

It is, therefore, desirable to consider a measure of extremal dependence for levels $t < 1$. One reason for this is the possibility of dependence for $t < 1$ while we observe asymptotic independence; additionally, such a measure will be helpful for empirical analyses, as these can only consider exceedance probabilities for $t < 1$. This latter point will be discussed in more detail in the context of nonparametric tail dependence estimation.

Rewriting Equation (2.63) as

$$\lambda_U = \lim_{t \to 1^-} \left(2 - \frac{1 - C(t,t)}{1-t} \right) = \lim_{t \to 1^-} \lambda_U(t) \; , \tag{2.76}$$

we obtain a quantity $\lambda_U(t)$ which can, in principle, convey information about extremal dependence at $t < 1$. As pointed out by Coles et al. (1999), the limit in Equation (2.76) can alternatively be expressed as

$$\lambda_U \;\; = \;\; \lim_{t \to 1^-} \left(2 - \frac{\ln(C(t,t))}{\ln(t)} \right) = \lim_{t \to 1^-} \chi(t) =: \chi \; , \tag{2.77}$$

which provides us with a second quantity, $\chi(t)$, which may be used as quantile–dependent measure of extremal dependence. Its sign depends on whether the variables under consideration are positively or negatively associated at quantile t.

In the case of asymptotic independence ($\lambda_U = \chi = 0$), Coles et al. (1999) suggest a measure whose limit conveys information about the strength of extremal dependence. They define

$$\bar{\chi} = \lim_{t \to 1^-} = \frac{2 \ln \left(\Pr[U_1 > t] \right)}{\ln(\Pr[U_1 > t, U_2 > t])} - 1 = \lim_{t \to 1^-} \frac{2 \ln(1-t)}{\ln(C^*(t,t))} - 1 = \lim_{t \to 1^-} \bar{\chi}(t) \tag{2.78}$$

where $C^*(t,t) = \Pr[U_1 > t, U_2 > t]$. If the random variables are tail dependent, $\bar{\chi} = 1$ must hold; in general, $-1 \leq \bar{\chi} \leq 1$. We therefore have obtained a second measure which can convey information about the presence of asymptotic dependence. Furthermore, in the case of tail independence, $\bar{\chi}$ provides a measure increasing with the strength of extremal dependence for $t < 1$. For example, a Gaussian copula with $\rho < 1$ implies $\lambda_U = \chi = 0$; however, $\bar{\chi} = \rho$ (Galambos, 1987). Table 2.4 summarizes the implications of the two asymptotic measures. For $\lambda_U = 0$, there is asymptotic (tail) independence, irrespective of the value of $\bar{\chi}$.[12] However, asymptotic dependence requires not only that $\lambda_U > 0$, but also that $\bar{\chi} = 1$. Still, for the case of asymptotic

[12] In particular, is is possible that $\bar{\chi} = 1$ for asymptotic independence.

	Asymptotic Independence	Asymptotic Dependence
$\lambda_U = \chi$	0	(0,1]
$\bar{\chi}$	[-1,1]	1

Table 2.4: Asymptotic (In)Dependence

(tail) independence, the value of $\bar{\chi}$ is not meaningless; but rather than considering the limit $t \to 1$, it refers to lower confidence levels. Those values of t are the relevant confidence regions in empirical applications, and the lower t, the higher the number of observations on which nonparametric estimates of quantile dependence can be based.

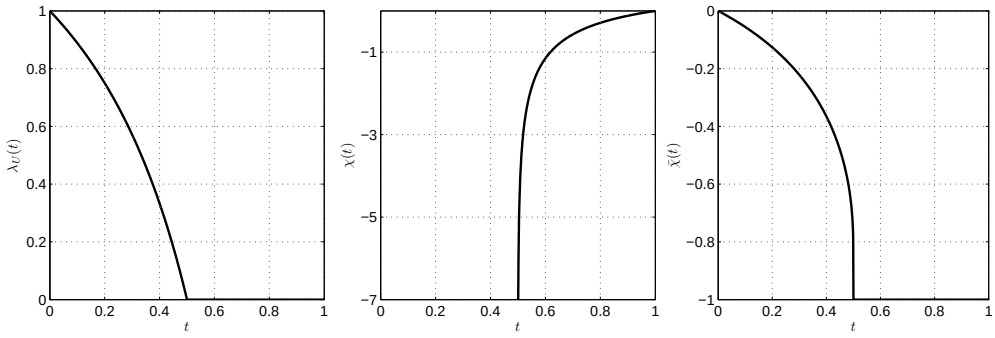


Figure 2.10: Quantities $\lambda_U(t)$, $\chi(t)$ and $\bar{\chi}(t)$ for the Countermonotonicity copula.

In order to illustrate the behavior of $\lambda_U(t)$, $\chi(t)$ and $\bar{\chi}(t)$ for different levels of t, we consider some examples. For the Fréchet upper bound, all three quantities λ_U, χ and $\bar{\chi}$ are equal to one for all levels of t and thus for any quantile considered. This implies perfect tail dependence as well as perfect dependence at all levels of t, corresponding to the notion of comonotonicity. Similarly, the independence copula leads to $\lambda_U = \chi = \bar{\chi} = 0$ for all t. However, the countermonotonicity copula exhibits the patterns shown in Figure 2.10. We see that its behavior changes at $t = 0.5$, which is due to the maximum expression taken in the calculation of C_u. The upper tail dependence estimators λ_U and χ converge towards zero: Although (negative) dependence is perfect, it cannot be captured due to its structure. However, $\bar{\chi} = -1$ hints at perfect negative association.

Figure 2.11 shows examples for different Archimedean and elliptical copulas as well as different levels of t and values of the dependence parameter. For the Gaussian copula, we see that $\lambda_U(t)$ and $\bar{\chi}(t)$ increase with t; by definition, their limits for $t \to 0$ coincide. Except for a correlation of $\rho = 1$, this limit equals zero due to tail independence. Correspondingly, $\bar{\chi}$ equals one only for perfect positive correlation. These graphs illustrate that for $\rho < 1$, there is extremal dependence at $t < 1$, but tail independence.

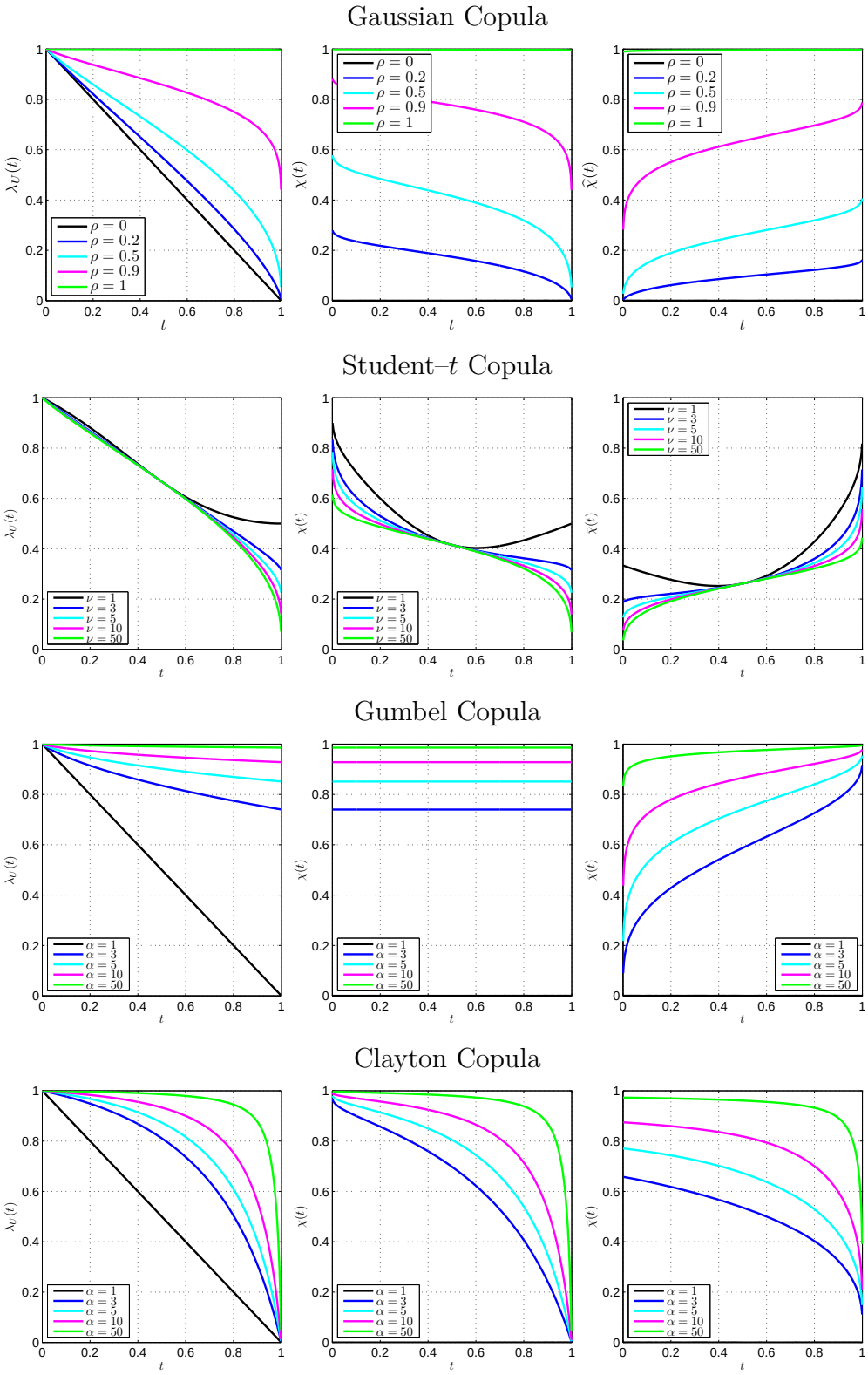


Figure 2.11: Quantities $\lambda_U(t)$, $\chi(t)$ and $\bar{\chi}(t)$ for Parametric Copulas

For the Student–t copula, the differences to the Gaussian are more pronounced, the lower the degrees–of–freedom parameter, ν. Here, $\bar{\chi}$ converges to one in all cases, corresponding to tail dependence. As can be seen from λ_U and χ, the tail dependence is the higher, the lower ν.
Similarly, we confirm upper tail dependence for the Gumbel copula[13], the strength of which can be assessed from λ_U and χ; in contrast to this, the Clayton copula does not exhibit upper tail dependence.

Estimation of Tail Dependence

Assuming that the underlying copula is in the domain of attraction of an extreme value copula, relationships from multivariate extreme value theory may be used in order to derive tail dependence estimates. For example, Frahm et al. (2005) suggest the following estimator of λ_U, which is motivated by Capéraà et al. (1997):

$$\widehat{\lambda}_U^{\mathrm{CFG}} = 2 - 2\exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln\left\{\sqrt{\ln(u_{1,i}^{-1})\ln(u_{2,i}^{-1})}/\ln\left((\max(u_{1,i},u_{2,i})^2)^{-1}\right)\right\}\right] . \quad (2.79)$$

Furthermore, the relationships between Kendall's τ and the upper tail dependence coefficient as given by tables 2.2 and 2.3 may be used to infer estimates for λ_U. However, we are interested in a purely nonparametric estimation of tail dependence, in order to gain insights independent of and additional to parametric copula inference.

The quantities $\lambda_U(t)$, $\chi(t)$ and $\bar{\chi}(t)$ can be used for estimation of tail dependence. Inserting the empirical copula from Equation (2.60), this can be achieved without any parametric assumption with respect to the underlying copula. The estimators thereby obtained are

$$\widehat{\lambda}_U(t) = 2 - \frac{\ln \widehat{C}_n(t,t)}{\ln(t)} , \quad (2.80a)$$

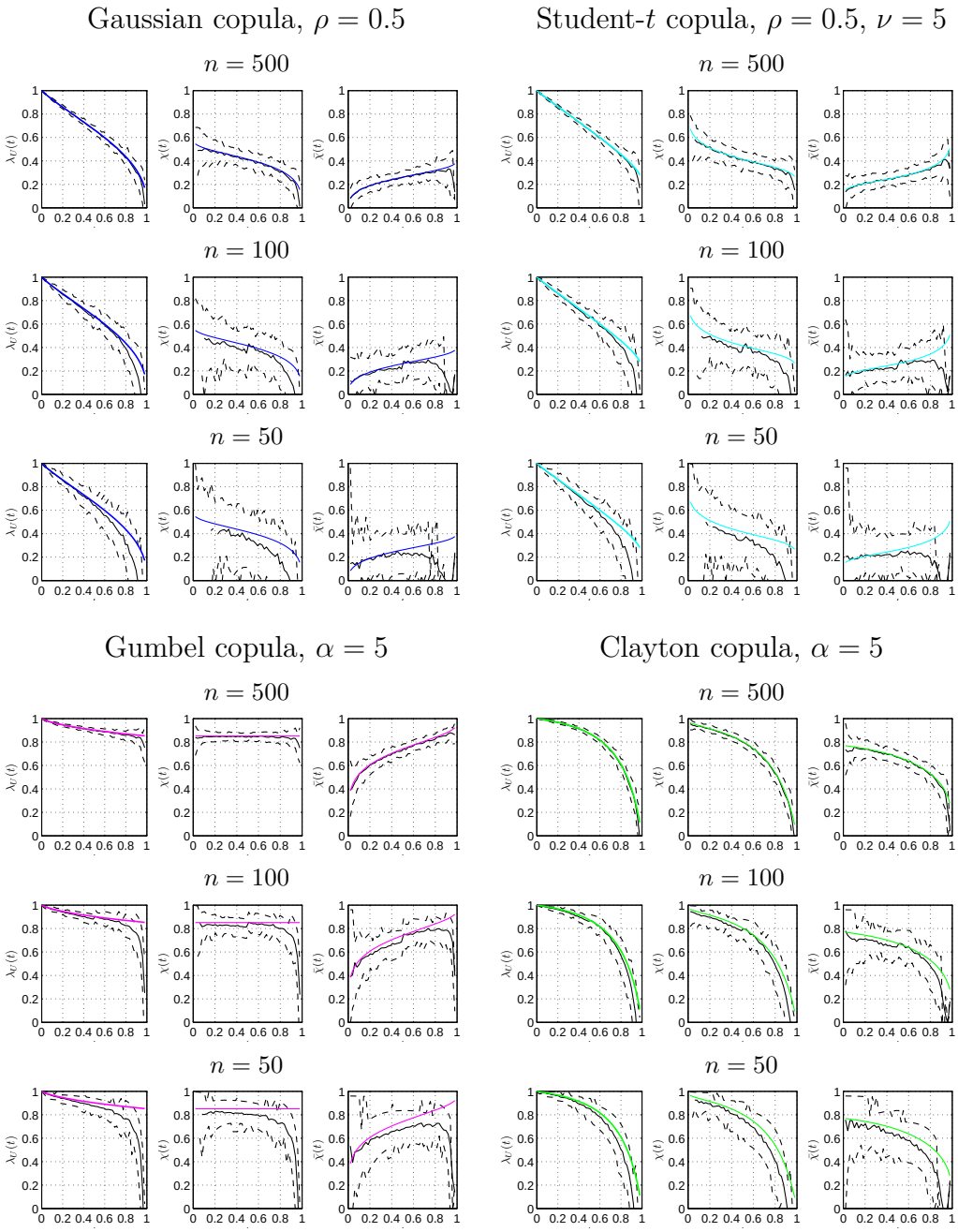
$$\widehat{\chi}(t) = 2 - \frac{1-\widehat{C}_n(t,t)}{1-t} , \quad (2.80b)$$

$$\widehat{\bar{\chi}}(t) = \frac{2\ln(1-t)}{\ln\left(\widehat{C^*}_n(t,t)\right)} - 1 . \quad (2.80c)$$

We analyze the effects of different sample sizes on estimation results in a small simulation study. That is, for each of four copulas with given parameter value, we generate n_{obs} observations. We then estimate tail dependence via Equations (2.80a) to (2.80c). This is done 100 times for each level of t. Figure 2.12 shows the resulting means as

[13] In fact, $\chi(t)$ is constant in t for all bivariate extreme value distributions.

well as the 5% and 95% quantiles of these estimates. For easier comparison, the corresponding true extremal dependence functions $\lambda_U(t)$, $\chi(t)$ and $\bar{\chi}(t)$ are included as colored lines.



Means (solid), 5% and 95% quantiles (dashed) of 100 estimates of $\lambda_U(t)$, $\chi(t)$ and $\bar{\chi}(t)$. The colored lines refer to the underlying true quantities.

Figure 2.12: Nonparametric Tail Dependence Estimation

We see that based on $n_{obs} = 500$ observations, the shapes of the extremal dependence functions are captured satisfactorily. Most importantly, in case of tail dependence, $\widehat{\lambda}_U(t)$ and $\widehat{\chi}(t)$ are close to λ_U and $\widehat{\bar{\chi}}(t)$ approaches one for high levels of t. It is, thus, possible to infer tail dependence. However, this is not the case for smaller sample sizes. In fact, the differences between true and estimated quantities increase with t, as a higher t implies fewer observations on which the estimation can be based. But this means that estimation is particularly difficult in the most important area of the distribution. For example, considering the Gumbel copula with $\alpha = 5$ (implying an substantial upper tail dependence of $\lambda_U = 0.8513$), an estimation based on $n_{obs} = 50$ observations could well lead to the conclusion of asymptotic independence, as all three estimators decrease towards zero near $t = 1$. This is a general problem which has to be faced when estimating exceedance probabilities and cannot be circumvented.

Tail dependence is typically estimated on either one or two of the quantities $\lambda_U(t)$, χt and $\bar{\chi}(t)$. For example, Frahm et al. (2005) compare the behavior of $\lambda_U(t)$ and $\chi(t)$, while Dupuis and Jones (2006) consider $\chi(t)$ and $\bar{\chi}(t)$. We argue that it is not necessary to take a decision as to which estimator is the best; rather, the behavior of all three quantities should be examined. If the underlying distribution is characterized by tail dependence, both $\lambda_U(t)$ and $\chi(t)$ should converge towards one, so that assessing both estimators may give more insight with respect to the question of (absence of) tail dependence than when only considering one quantity. Similarly, $\widehat{\bar{\chi}}$ should be checked to converge towards one if tail dependence is inferred from a sample. Furthermore, as can be seen from Figure 2.11, different parametric copulas imply different shapes of the extremal dependence functions; this fact will be used in the context of assessing the goodness–of–fit of parametric copulas.

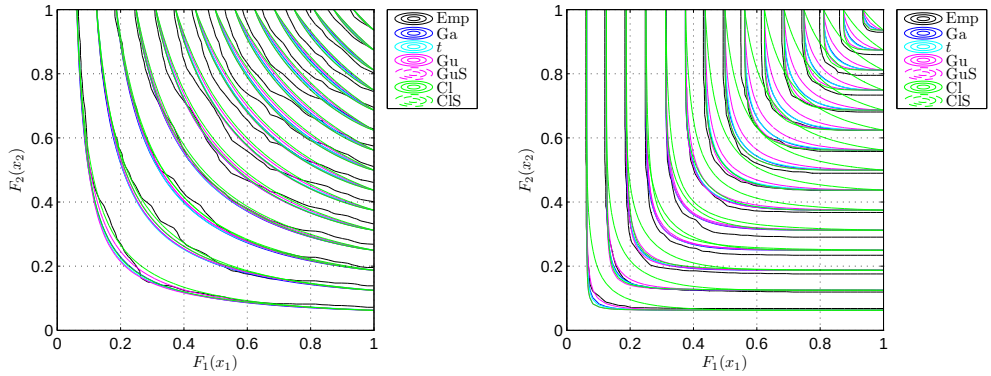
2.3.3 Goodness of Fit

Once the parameters of different copulas have been estimated, the question of how to compare their fit to the observed data has to be answered. A first approach consists of purely graphical methods.

Example 2.1. *In the goodness–of–fit analyses presented in the remainder of this chapter, we will use two scenarios which differ in the strength of (tail) dependence. Both will use data simulated under the assumption of a Gumbel copula, in order to assess the effects of asymmetric, nonlinear dependence.*

1. ***Moderate Dependence.*** *Here, we choose an upper tail dependence coefficient of $\lambda_U = 0.2$. This implies that the parameter of the Gumbel copula is equal to $\alpha = 1.179$, and a (Kendall) rank correlation of $\tau = 0.1518$.*

2. ***Strong Dependence.*** *Setting* $\lambda_U = 0.8$ *leads to* $\alpha = 3.802$ *and* $\tau = 0.737$.



(i) moderate dependence (ii) strong dependence

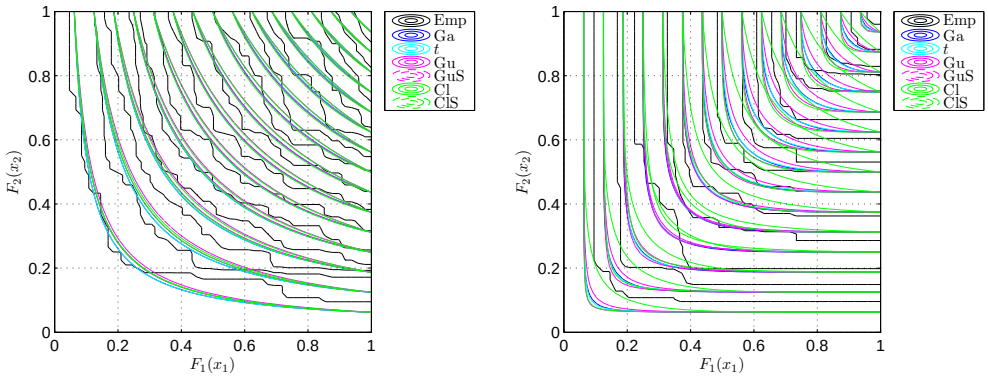
Figure 2.13: The graphical approach to copula goodness–of–fit testing: $n = 1{,}000$.

Figure 2.13 shows contour plots of the empirical copula against the contours of the copulas fitted via Maximum Likelihood. We observe that the latter are closer in the case of moderate dependence, which makes perfectly sense since one of the main distinguishing dependence structures—tail dependence—is only weak. In contrast to this, for strong dependence, the differences among the shapes given some level of $C(\cdot, \cdot)$ *are clearly visible. This is remarkably true for the Clayton survival copula, which is—due to its implication of absence of higher and presence of lower tail dependence—far from the empirical copula. We thus observe an effect which will be encountered several times in the following: The stronger the dependence, the better the discrimination among copulas. Figure 2.13 was based on* $n = 1{,}000$ *bivariate realizations. A realistic sample size in risk management applications will be typically lower, reducing the usefulness of such graphical (and, in general, all goodness–of–fit) methods. In Figure 2.14, the empirical copula is characterized by kinks. The shapes of the contours—and, thus, the distance of the theoretical contours to the empirical ones—are thus largely driven by single observations.*

As a different means towards evaluation of copula goodness–of–fit, information criteria such as the Akaike (AIC) and Bayesian (BIC) may be employed. With q estimated parameters and sample size T, these are given by

$$\begin{aligned} AIC &= -2\ln(\ell(\theta)) + 2q \\ BIC &= -2\ln(\ell(\theta)) + q\ln(T)\,, \end{aligned} \tag{2.81}$$

where $\ell(\theta)$ denotes the maximized log–likelihood. One chooses that model which yields the lowest value for AIC and/or BIC. However, these information criteria do not give



(i) moderate dependence (ii) strong dependence

Figure 2.14: The graphical approach to copula goodness–of–fit testing: $n = 100$.

any hint with respect to the size and power of the decision rule; therefore, goodness–of–fit tests are preferred.

Typically, the hypotheses we wish to test are

$$\mathcal{H}_0 : C \in \mathcal{C} = \{C_\theta; \theta \in \Theta\} \quad \text{vs.} \quad \mathcal{H}_1 : C \notin \mathcal{C} = \{C_\theta; \theta \in \Theta\} \tag{2.82}$$

That is, we have in fact a composite hypothesis which contains the copula family as well as its parameter value.
If we are able to express the hypotheses in terms of a univariate random variable, usual goodness–of–fit tests such as the Cramér–van–Mises (CvM, see, for example, Pearson and Stephens, 1962), Kolmogorov–Smirnov (KS, Kolmogorov, 1933; Smirnov, 1939) or the Anderson–Darling (AD, Anderson and Darling, 1952) test. Generally speaking, they are based on the distance between the empirical distribution ($\widehat{F}(x)$) and the theoretical one ($F(x)$). The CvM statistic is given by

$$\text{CvM} = n \int_0^1 (\widehat{F}(x) - F(x))^2 \mathrm{d}F(x) \; . \tag{2.83}$$

The KS statistic is obtained from

$$\text{KS} = \max_x \left(|\widehat{F}(x) - F(x)| \right) \; , \tag{2.84}$$

and the AD test can be written as

$$\text{AD} = \max_x \frac{|\widehat{F}(x) - F(x)|}{\sqrt{F(x)(1 - F(x))}} \; . \tag{2.85}$$

The AD test statistic puts more emphasis on and is thus more accurate in the tails of $\widehat{F}(x)$ and $F(x)$. Correspondingly, the KS test is more sensitive towards changes in the

center of the distribution. Additionally, the AD test is characterized by a high power against various alternatives (D'Agostino and Stephens, 1986).

All three statistics can thus be expressed as continuous functionals of the empirical process

$$\sqrt{n}\left(\widehat{F}(x) - F(x)\right) , \tag{2.86}$$

where $F(x)$, and thus the asymptotic distribution of the test statistic, depends on the–typically unknown–parameter θ. In other words, whenever we do not know all parameters, we may not resort to their theoretical distributions.

In the following, we first present three prominent types of goodness–of–fit tests for copulas. The first one is an adaptation of the well–known univariate χ^2 test to multivariate settings. Through the calculation of the test statistic, the multivariate problem is transformed to a univariate one. Similarly, the tests outlined in Sections 2.3.3.2 and 2.3.3.3 are based on differences between empirical and theoretical univariate quantities. They are reviewed—among a variety of others—and compared by Berg (2009), whose results lead to the choice of tests considered here. They will be discussed in the corresponding sections. Finally, Section 2.3.3.4 suggests a new criterion for the choice among different copula families which is inspired by nonparametric tail dependence estimation.

2.3.3.1 χ^2 Tests

A goodness–of–fit test for copulas based on the χ^2 distribution is used by Genest and Rivest (1993) and Hu (2006) in order to assess the fit of different copula models. A more detailed account on this approach can be found in Dobric and Schmid (2005); the latter paper modifies the hypothesis to be tested such that only the family of copula is formulated as null hypothesis.[14]

In a first step, the sample space is divided into classes, in order to be able to build a contingency table used for testing. For example, Dobric and Schmid (2005) partition the unit square into $r \times s$ rectangles of equal area $1/(rs)$, while Genest and Rivest (1993) and Hu (2006) construct the (upper) boundaries for a sample of size n and k^2 classes (in other words, $r = s = k$) as the order statistics at position $n * j/k$, $j = 1, \ldots, k$.[15] Whatever the choice for the construction of the bounds, its result are B_{ij} cells partitioning the sample space. Given an estimate $\widehat{\theta}$ for the copula parameter,

[14] The parameter value of the copula is then chosen such that the χ^2 test statistic is minimized.

[15] The setup of Genest and Rivest (1993) and Hu (2006) leads to an equal number of observations per row and column, that is, equal marginal frequencies.

one can then calculate the theoretical joint probability of cell B_{ij}

$$p_{ij}(\widehat{\theta}) = \Pr\left[\boldsymbol{X} \in B_{ij}|\widehat{\theta}\right] \ . \tag{2.87}$$

Denoting the number of observations in cell B_{ij} by N_{ij}, the test statistic is given by

$$\chi^2(\widehat{\theta}) = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(N_{ij} - np_{ij}(\widehat{\theta})\right)^2}{np_{ij}(\widehat{\theta})} \tag{2.88}$$

and follows a χ^2 distribution with $\text{df} = (r-1)(s-1) - d$ degrees of freedom, where d denotes the number of parameters estimated.[16]

Example 2.2. *We draw bivariate realizations from the Gumbel copula, the parameter value depending on the scenario used. For simplicity, the margins are assumed to be standard normal. In order to apply the χ^2 test, we have to construct a contingency table. As in Genest and Rivest (1993) and Hu (2006), we choose $k = 7$ cells, whose boundaries are given by the order statistics at position $n * j/k = 1000 * j/7$ for $j = 1, \ldots, 6$. For the cells constructed in this way, we evaluate the number of observations they contain.*

1. ***Moderate Dependence.*** *Table 2.5 shows the contingency table of these $n =$* 1,000 *bivariate observations.*

	X_2						
	29	29	19	18	20	16	12
	23	21	27	20	19	14	19
	20	24	21	20	22	23	13
X_1	19	23	21	22	16	23	18
	18	20	20	25	22	19	19
	20	13	18	24	27	23	18
	14	13	17	13	17	25	44

Table 2.5: χ^2 Test: Number of Observations, $\lambda_U = 0.2$

As is to be expected for a positive dependence setup, the frequencies along the main diagonal tend to be higher than those in the other cells. However, the distribution of the $n =$ 1,000 *observations across cells appears quite even. The p–values obtained from this setup are shown Table 2.6. We see that none of the null hypotheses can be rejected at any reasonable confidence level.*

[16] Intuitively, the number of rows and columns has to be reduced due to the fact that the last row (column) is given and cannot be chosen freely.

Copula	Ga	t	Gu	GuS	Cl	ClS
$n = 1{,}000$	0.8028	0.9448	0.9899	0.4198	0.1242	0.9717
$n = 500$	0.3599	0.3888	0.5534	0.1292	0.0445	0.6086

Table 2.6: χ^2 Test: p–values, $\lambda_U = 0.2$.

It has to be kept in mind that this situation, which leaves us without any clear–cut hint at the most appropriate copula, was created based on $n = 1{,}000$ observations, which is a very comfortable situation in a risk–management framework. We can easily construct examples where the highest p–value belongs to a copula other than that used for generating the data: The second row of Table 2.6, which is based on the same parameter values as before, but a sample size of $n = 500$, leads to a decision in favor of the Clayton survival copula.

2. ***Strong Dependence.***

The distribution of observations across the $k^2 = 49$ cells shows now clearly the tendency of clustering along the main diagonal. However, this clearer dependence pattern creates a new problem in conducting the test.

	X_2						
	98	31	11	**2**	**1**	**0**	**0**
	30	57	36	14	**4**	**2**	**0**
	11	40	50	29	11	**1**	**1**
X_1	**4**	11	34	50	31	10	**2**
	0	**4**	9	39	61	30	**0**
	0	**0**	**3**	7	33	73	27
	0	**0**	**0**	**1**	**2**	27	113

Table 2.7: χ^2 Test: Number of Observations, $\lambda_U = 0.8$

As can be seen from Table 2.7, the clustering of observations leads to several (boldface) cells containing only few (≤ 4) or no observations. The common remedy to this problem consists of pooling together cells until all of them contain at least 5 observations. Obviously, this pooling can be done in different ways; we arbitrarily choose among these possibilities. That is, we create four "new" cells by pooling cells according to their colors; for example, one new cell sums up observations of B_{14}, B_{15} and B_{25}. Next, we pool the neighboring new cells, so that we obtain only

2 new cells—one below, one above the main diagonal. Finally, we also unite these two, so that we use only one new cell. Table 2.8 contains the p–values thereby obtained (we exclude those copulas for which the p–values are zero for all pooling schemes); the second row corresponds to the case where pooling is neglected and all original cells are used. We see that without any pooling, the p–values of the

	Number of pooled cells			
	0	1	2	4
Gaussian	0.0000	0.2417	0.2933	0.1694
Student–t	0.2720	0.2415	0.2221	0.2449
Gumbel	0.3091	0.8164	0.8679	0.8985

Table 2.8: χ^2 Test: p–values, $\lambda_U = 0.8$

Student-t and the Gumbel copula are close, while all other null copulas can be rejected at any confidence level. However, if we start pooling, using only one new cell, the Gaussian copula cannot be rejected any longer, while the Student–t and Gumbel p–values differ considerably. For more granular pooling schemes, results clearly hint at the Gumbel copula.

As is to be expected, this test performs better, the stronger the dependence among the data. However, such a dependence structure may call for a pooling of cells, which, in turn, has a influence on testing. This latter phenomenon can be reproduced for different simulation setups: Depending on the classification and pooling scheme chosen, the choice among null copulas may differ. Therefore, this sort of goodness–of–fit test contains a subjective element.

2.3.3.2 Tests Based on the Rosenblatt Transform

A class of goodness–of–fit tests for copulas relies on the transform of Rosenblatt (1952). Given a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with absolutely continuous joint distribution function $F_{\boldsymbol{X}}$ and margins $F_i(x_i)$, it is given by

$$\begin{aligned} T(x_1) &= \Pr[X_1 \leq x_1] = F_1(x_1) \\ T(x_2) &= \Pr[X_2 \leq x_2 | X_1 = x_1] = F_{2|1}(x_2|x_1) \\ \ldots &\quad \ldots \\ T(x_d) &= \Pr[X_d \leq x_d | X_1 = x_1, \ldots, X_{d-1} = x_{d-1}] = F_{d|1,\ldots,d-1}(x_d|x_1, \ldots, x_{d-1}) \end{aligned} \tag{2.89}$$

The random variables obtained from this transformation, $Z_i = T_i(x_i)$, are uniformly and independently distributed on $[0, 1]^d$.

This idea can be used to test for the goodness–of–fit of parametric copulas by expressing the transformation in terms of (conditional) copulas. Denoting by $C_i(u_1, \ldots, u_d)$ the joint i–marginal distribution of $\boldsymbol{U}$, that is,

$$C_i(u_1, \ldots, u_i) = C(u_1, \ldots, u_i, 1, \ldots, 1) \; , \tag{2.90}$$

the conditional distribution of U_i given $U_1, \ldots, U_{i-1}$ is given by

$$C_i(u_i|u_1, \ldots, u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, \ldots, u_i)/\partial u_1 \cdots u_{i-1}}{\partial^{i-1} C_{i-1}(u_1, \ldots, u_i)/\partial u_1 \cdots u_{i-1}} \tag{2.91}$$

for $i = 2, \ldots, d$. The transformed random variables obtained from (2.89) can thus alternatively be expressed as

$$Z_i = C_i(F_i(x_i)|F_1(x_1), \ldots, F_{i-1}(x_{i-1})) \; . \tag{2.92}$$

If $C_i(\cdot)$ represents the ("true")underlying (conditional) copula of $\boldsymbol{X}$, we should, accordingly, find that $Z_i \overset{\text{iid}}{\sim} \text{Unif}(0, 1)$. Applying the probability–integral transformation, we may then use any distribution for testing. Typically, one applies the inverse of the standard normal distribution, so that $\Phi^{-1}(z_i)$, which must be standard normally distributed if C_i is correctly representing the data's dependency structure. A test statistic can then be derived from $S = \sum_{i=1}^{d} (\Phi^{-1}(z_i))^2$, which accordingly follows a χ^2_d distribution. For the two–dimensional case, to which we are restricting attention, the strategy of goodness–of–fit testing via the Rosenblatt transform thus boils down to testing whether

$$S(x_1, x_2) = \left[\Phi^{-1}(F_1(x_1))\right]^2 + \left[\Phi^{-1}(C_2(F_2(x_2)|F_1(x_1)))\right]^2 \tag{2.93}$$

follows a χ^2_2 distribution, where the empirical distribution functions are inserted for $F_i(x_i)$, and $C_2(\cdot, \cdot)$ is evaluated from Equation (2.91). This latter quantity can easily be computed for Archimedean copulas. For the implicit Gaussian and Student–t copulas, one may use the fact that the conditional copula can be expressed as (see, for example, Cherubini et al., 2004)

$$C_2^{\text{Ga}}(u_2|u_1) = \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\,\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \tag{2.94}$$

and

$$C_2^{\text{t}}(u_2|u_1) = t_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu + t_\nu^{-1}(u_1)^2}}\,\frac{t_\nu^{-1}(u_2) - \rho\, t_\nu^{-1}(u_1)}{\sqrt{1-\rho^2}}\right) \; , \tag{2.95}$$

where $t_\nu^{-1}(\cdot)$ refers to the cumulative distribution function of the Student–t distribution with ν degrees of freedom.

This approach to goodness–of–fit testing was first applied in a risk management framework by Berkowitz (2001), followed by Breymann et al. (2003) and Malevergne and

Sornette (2003); the latter considering only the null hypothesis of a Gaussian copula. However, it is not as straightforward as it may seem at first sight. Firstly, observations are transformed via their empirical distribution functions; this use of ranks introduces dependence, and thus Z_i is only close to iid. This has been pointed out, for example, by Dobric and Schmid (2007). Secondly, the calculation of S involves *estimated* parameter values, on which the distribution of test statistics depend (see, for example, Babu and Rao, 2004; Genest and Rémillard, 2008). The use of estimated parameter values and standard critical values can thus only be justified—at most—asymptotically. To tackle this issue, p–values for such a goodness–of–fit test need to be obtained via bootstrapping (Efron and Tibshirani, 1986).[17] The procedure is sketched in Table 2.9.

1. Calculate empirical cdf of $\boldsymbol{x_1} = (x_{11}, \ldots, x_{1n})', \boldsymbol{x_2} = (x_{21}, \ldots, x_{2n})'$
2. Estimate parameters θ
3. Using $u_i = \widehat{F}_i(x_i)$ from 1. and $\widehat{\theta}$ from 2., calculate S and the corresponding Anderson–Darling test statistic AD
4. **Parametric Bootstrap**: Repeat B times the following steps.
   - Generate a sample of size n from the null hypothesis copula, $\boldsymbol{x_1}^0, \boldsymbol{x_2}^0$, using the estimated parameter value $\widehat{\theta}$
   - Calculate empirical cdf of $\boldsymbol{x_1}^0, \boldsymbol{x_2}^0$
   - Estimate parameters θ^0
   - Calculate S^0 and corresponding Anderson–Darling test statistic AD_b^0
5. Calculate the approximate p–value from $\hat{p} = 1/(B+1) \sum_{b=1}^{B} \mathcal{I}_{AD_b^0 > AD}$

Table 2.9: GoF testing Based on the Rosenblatt Transform.

This approach to goodness–of–fit testing for copulas has been generalized by Berg and Bakken (2005). They consider a more general version of Equation 2.93, that is, the quantity

$$\widetilde{S} = \sum_{i=1}^{d} \Gamma(z_i, \boldsymbol{\alpha}) \tag{2.96}$$

[17] This need for a bootstrap is neglected by Breymann et al. (2003), but taken into account by Malevergne and Sornette (2003).

for some weighting function Γ with weight parameters $\boldsymbol{\alpha}$. Using $\gamma(z_i; \boldsymbol{\alpha}) = \Phi^{-1}(z_i)^2$ as in Breymann et al. (2003) thus represents a special case. Berg and Bakken (2005) show their modification to be consistent. But this comes at a cost: Unlike the case of Equation (2.93), the distribution of $\widehat{S}$ is unknown and must be determined via a double bootstrap procedure.

Replacing the AD statistic with the more general continuous functional of the empirical process (2.86), the parametric bootstrap of Step 4 fits into the bootstrap procedure suggested and described by Stute et al. (1993) and Henze (1996) in the univariate case and extended towards multivariate setups by Genest and Rémillard (2008). The latter authors prove that this procedure yields a valid approximation to the null distribution.

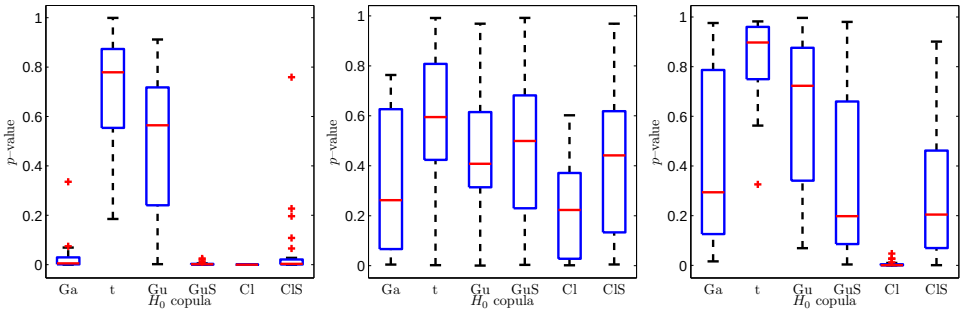
Example 2.3. *We again apply the test to artificial datasets consisting of bivariate observations from a Gumbel copula. The corresponding boxplots are shown in Figure 2.15. First, we use* $n = 1{,}000$ *observations and a copula parameter* $\alpha = 3.8$*. The number of replications in the bootstrap is set to* $B = 10{,}000$*.*[18] *Repeating* 25 *times, we obtain 25 different p-values for each null hypothesis copula.*

While for the Gaussian, Gumbel Survival, Clayton and Clayton survival copulas, the interquartile range is clearly below $p = 0.1$*, this is not true for the Student–t copula. It thus seems that the test cannot distinguish between symmetric and asymmetric dependence structures satisfactorily.*

If we reduce the upper tail dependence to $\lambda_U = 0.2$ *(*$\alpha = 1.179$*), the results change substantially. This is shown in the middle of Figure 2.15, where none of the null hypothesis copulas can be rejected any more and the interquartile ranges are much wider than in the case of strong dependence. Similarly, reducing the sample to only* $n = 100$ *observations (right part of Figure 2.15) leads to results where only one copula (the Clayton) can be rejected at a confidence level below 10%.*

We conclude that this type of copula goodness–of–fit test may be useful for large samples and in cases of substantial dependence; however, it may fail to detect asymmetries and its power may be weak in other situations. These findings agree with those of Berg (2009), who finds the Rosenblatt–based tests to perform poorly for a large variety of copulas.

[18] This number of replications corresponds to the settings of Malevergne and Sornette (2003) and Berg (2009) who find it to be sufficient in order to obtain reliable results.



$\lambda_U = 0.8, n = 1{,}000$ $\lambda_U = 0.2, n = 1{,}000$ $\lambda_U = 0.8, n = 100$

Figure 2.15: Simulation of p–values Based on the Rosenblatt Transform

2.3.3.3 Tests Based on the Empirical Copula

In the following, we present two tests which are found to perform especially well in a variety of setups by Berg (2009).[19] Both are based on the distance between the empirical and the null copula.

The Genest and Rémillard (2008) approach ($\mathcal{A}_2$)

This approach to goodness–of–fit testing reduces the dimensionality via Deheuvel's empirical copula from Equation (2.60), calculated from the empirical distribution functions. That is, in the two–dimensional setting,

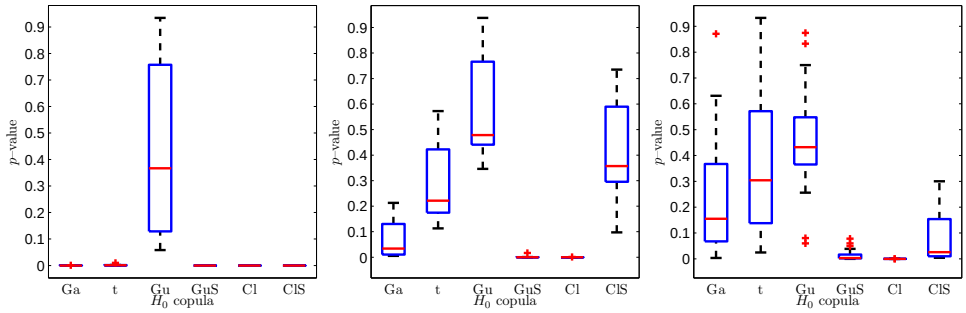
$$\widehat{C}(u_1, u_2) = \frac{1}{n+1} \sum_{j=1}^{n} \mathcal{I}_{(-\infty,u_1]} \left(\widehat{F}(x_1)\right) \mathcal{I}_{(-\infty,u_2]} \left(\widehat{F}(x_2)\right) \ . \tag{2.97}$$

Comparing this quantity to the theoretical copula C can be done, for example, via a Cramér–van–Mises test (Genest et al., 2009):

$$\mathrm{CvM}_{\mathcal{A}_2} = n \int_{[0,1]^d} \left(\widehat{C}(\boldsymbol{u}) - C(\boldsymbol{u})\right)^2 \mathrm{d}\widehat{C}(\boldsymbol{u}) = \sum_{i=1}^{n} \left(\widehat{C}(\boldsymbol{u}) - C(\boldsymbol{u})\right)^2 \ . \tag{2.98}$$

Example 2.4. *Considering the same setup as before, we see that the test performs clearly better than the one based on the Rosenblatt transform in the case of strong dependency and a large sample. That is, for all null hypotheses except the Gumbel, the p–values are virtually zero in all cases. Reducing the dependence ($\lambda_U = 0.2$) or the sample size ($n = 100$) leads to an increase in p–values; only for those copulas which imply absence of upper and presence of lower tail dependence (the Gumbel survival and Clayton copula), they remain zero.*

[19] Adopting from Berg (2009), we call these two approaches $\mathcal{A}_2$ and $\mathcal{A}_4$.



$\lambda_U = 0.8, n = 1{,}000$ $\lambda_U = 0.2, n = 1{,}000$ $\lambda_U = 0.8, n = 100$

Figure 2.16: Simulation of p–values, Approach $\mathcal{A}_2$

The Genest and Rivest (1993) approach ($\mathcal{A}_4$)

Another approach is based on the distribution function of the copula,

$$K(v) = \Pr[C(\boldsymbol{u}) \leq v] \ . \tag{2.99}$$

It can be estimated by the empirical distribution function of the empirical copula, that is,

$$\widehat{K}(v) = \frac{1}{n+1} \sum_{j=1}^{n} \mathcal{I}_{[0,v]} \left(\widehat{C}(\boldsymbol{u}) \right) \ . \tag{2.100}$$

Kendall's process, which was introduced by Genest and Rivest (1993), is then given by the empirical process

$$\alpha(v) = \sqrt{n} \left(\widehat{K}(v) - K(v) \right) \ . \tag{2.101}$$

Its name derives from the fact that it is closely related to Kendall's τ.[20] For an Archimedean copula with generator $\phi(t)$, this distribution function of $V = C(u_1, u_2)$ can be written as

$$K(v) = v - \lambda(v) \qquad \text{with} \qquad \lambda(v) = \frac{\phi(v)}{\phi'(v)} \ . \tag{2.102}$$

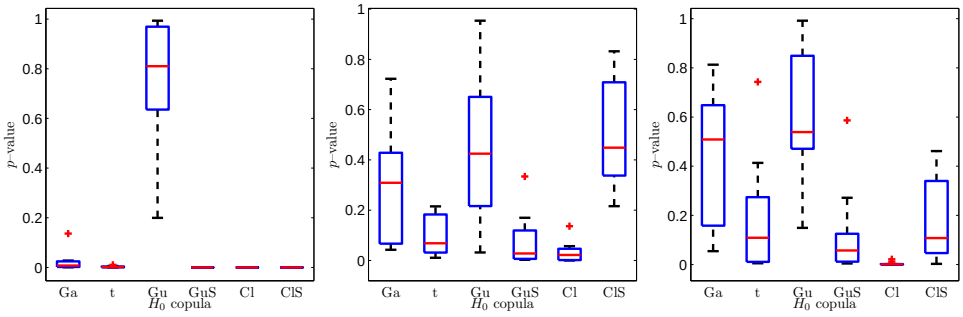
If $\phi/(u_1)/(\phi(u_1) + \phi(u_2)) \sim \text{Unif}(0,1)$ and independent of V, (2.102) implies that the underlying copula is Archimedean; see Genest and Rivest (1993) for a proof. Therefore, and as the relationship from Equation (2.102) can be solved for the copula's generator ϕ, estimation of $K(v)$ amounts to an estimation of the copula in an Archimedean setting. Comparing the empirical estimate (2.100) to its theoretical counterpart thus yields a valid approach for testing the goodness of fit. This can be done graphically, as

[20] That is, Kendall's τ is an affine transformation of the mean of the distribution function $K(\cdot)$. This relationship can be exploited for choosing among different Archimedean copula alternatives in a method–of–moments–style approach; see Genest and Rivest (1993).

demonstrated by Genest and Rivest (1993), or with one of the available test statistics. However, as we are not exclusively considering Archimedean copulas, and as $K_n(v)$ can not always be expressed in closed form, we have to use a double–bootstrap procedure in order to obtain an approximation to Kendall's process under the null hypothesis copula.[21] Again, we use a Cramér–van–Mises test statistic:

$$\mathrm{CvM}_{\mathcal{A}_4} = n \int_{[0,1]^d} \left(\widehat{K}(\boldsymbol{v}) - K(\boldsymbol{v})\right)^2 \mathrm{d}\widehat{K}(\boldsymbol{u}) = \sum_{i=1}^{n} \left(\widehat{K}(\boldsymbol{u}) - K_{\widehat{\theta}}(\boldsymbol{u})\right)^2 . \tag{2.103}$$

Example 2.5. *Within the setup discussed before, results for the case of strong dependence and* $n = 1{,}000$ *again look very promising, as is seen in the left part of Figure 2.17.*



$\lambda_U = 0.8, n = 1{,}000$ $\lambda_U = 0.2, n = 1{,}000$ $\lambda_U = 0.8, n = 100$

Figure 2.17: Simulation of p–values, Approach $\mathcal{A}_4$

However, we observe again a substantial change of results in cases of only moderate dependence (middle) or a small sample (right). Here, the p–values for the Gumbel survival copula, implying exactly the opposite dependence structure, may even be near 20%. Only the Clayton copula hypothesis can be rejected.

2.3.3.4 Copula Goodness–of–Fit Based on Quantile Dependence

A goodness–of–fit criterion which has, to our knowledge, not been suggested so far links the nonparametric estimators of tail dependence to parametric copulas. As seen in Section 2.3.2, three different quantities—λ_U, χ and $\bar{\chi}$—can be used in order to assess quantile and tail dependence. Furthermore, different parametric copulas imply different shapes of these functions depending on the confidence level, t (see Figure 2.11). If we estimate λ_U, χ and $\bar{\chi}$ via Equations (2.80a) to (2.80c), we should find that the distance between estimates and theoretical quantities—given some estimate of the copula parameters—should be minimal for the "true" copula. This very basic idea

[21] Conditions for weak convergence of Kendall's process can be found in Barbe et al. (1996).

may serve as a criterion for assessing the goodness of fit of several parametric copulas: Given their parameter estimates, we need to compare the shapes obtained for $\widehat{\lambda}_U(t)$, $\widehat{\chi}(t)$ and $\widehat{\bar{\chi}}(t)$ to those implied by the parametric copulas, for different levels of t. This can be done visually; alternatively, we can use some distance measure.

Example 2.6. *We again draw $n = 1{,}000$ bivariate realizations from a Gumbel copula with parameter $\alpha = 3.8$. We estimate λ_U, χ and $\bar{\chi}$ for a range of values of $t \in (0, 1)$, as well as the parameters of our candidate copulas via Maximum Likelihood. The resulting graph of empirical and theoretical extremal dependence functions is shown in Figure 2.18.*

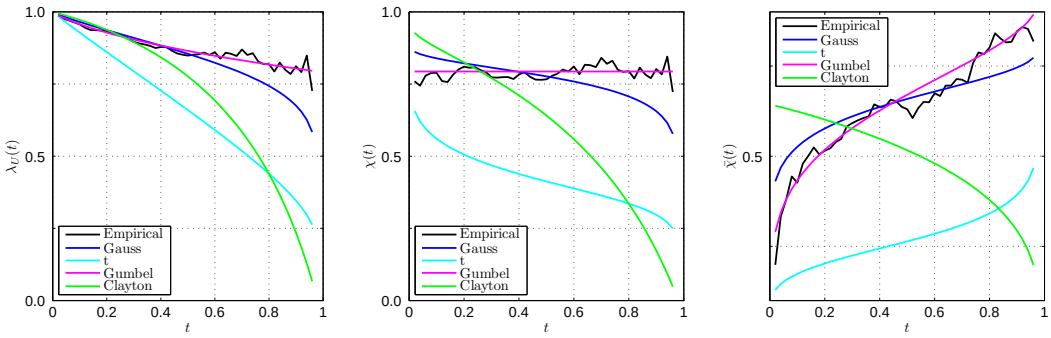
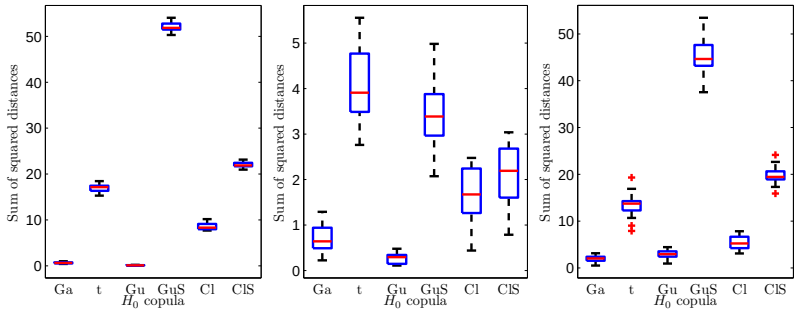


Figure 2.18: Empirical vs. Theoretical Extremal Dependence.

As expected, the empirically estimated functions are closest to the theoretical ones corresponding to a Gumbel copula with parameter $\widehat{\alpha} = 3.8143$. Therefore, a purely visual inspection clearly hints at the Gumbel copula. In order to additionally measure the distance between the functions, we define the criterion

$$d = \sum_{j=1}^{m} \left(\widehat{\lambda}_U(t_j) - \lambda_U(t_j)\right)^2 + \sum_{j=1}^{m} \left(\widehat{\chi}(t_j) - \chi(t_j)\right)^2 + \sum_{j=1}^{m} \left(\widehat{\bar{\chi}}(t_j) - \bar{\chi}(t_j)\right)^2 , \qquad (2.104)$$

consisting of the sum of squared distances between nonparametric and theoretical quantities. Running simulation and estimation 25 times, we obtain boxplots of the sums of squared distances. These are shown in Figure 2.19. We see that the Gaussian copula leads to distances very close to those of the ("true") Gumbel copula. This can be confirmed by inspection of Figure 2.18, where the extremal dependence functions implied by the Gaussian copula are closest to those of the Gumbel copula. However, the Student–t implied distances are substantially higher than those of the Gumbel, so that—different from the approaches presented so far—the Student–t can be easily distinguished from the Gumbel copula. For moderate dependence, the Gumbel criterion remains the smallest; for a small sample, the Gaussian copula implies the smallest value. However, the ordering of the copulas seems to be less affected by effects of small samples and moderate dependence than the goodness–of–fit approaches presented so far.



$\lambda_U = 0.8, n = 1{,}000$ $\lambda_U = 0.2, n = 1{,}000$ $\lambda_U = 0.8, n = 100$

Figure 2.19: Simulation of a New Goodness–of–Fit Criterion.

We conclude that nonparametric measures of extremal dependence can, in principle, be used to assess the goodness of fit of parametric copulas. This follows directly from visual inspection of Figure 2.18. A simple distance measure may help to find the parametric copula implying extremal behavior closest to the one empirically observed and nonparametrically estimated. A formal test remains to be developed and shown to yield valid results under various setups concerning null copulas and parameter values.

2.4 Measuring Dependent Risks

2.4.1 Risk Measures

One of the steps in the risk–management process is the measurement of risk. A quantification of existing risks is necessary in order to be able to establish limits, for example, for credit lines, or to set up capital buffers against potential adverse movements leading to large losses. We will mainly be concerned with the latter issue, that is, estimating risk capital for a given loss (or return) distribution. Other approaches to risk quantification which are not based on loss distributions, such as scenario–based measures, will not be considered here.

Using a distribution of past losses allows a consideration of aggregate loss across risk types and on an enterprise–wide perspective, because losses are a universal concept in all risk types. In this way, we can merge losses from, say, a portfolio of credit counterparties, with a portfolio of stocks to obtain an aggregate loss distribution. However, it has to be kept in mind that approaches relying purely on historical losses are, by their very nature, not able to predict movements beyond those observed in the past.[22]

[22]Using a popular illustration, one can say that such an approach amounts to driving a car while looking in the rear–view mirror.

Value–at–Risk (VaR)

The Value–at–Risk (VaR) at confidence level α of a distribution of losses, L, is given by

$$\mathrm{VaR}_\alpha = \inf\{\ell \in \mathbb{R} : \Pr[L > \ell] \leq 1 - \alpha\} = \inf\{\ell \in \mathbb{R} : F_L(\ell) \geq \alpha\} \,, \tag{2.105}$$

which is the α–quantile of the loss distribution. It thus denotes the maximum loss which will not be exceeded with a probability of at least $\alpha\%$. VaR is the most commonly used risk–measure due to its simplicity and the easiness with which its concept can be grasped. However, VaR suffers from drawbacks that should be known when applying it in order to derive risk–capital estimates. Firstly, it is not able to convey information about the distributional properties of those $(1 - \alpha)\%$ of losses exceeding it. This is illustrated in Figure 2.20, which shows two different loss histograms. Up to and including $\mathrm{VaR}_{0.95}$, they are identical; however, observed losses are higher beyond this point for the distribution on the right. An exceedance of VaR will thus tend to lead to higher losses and thus to a higher risk, which cannot be captured by VaR.

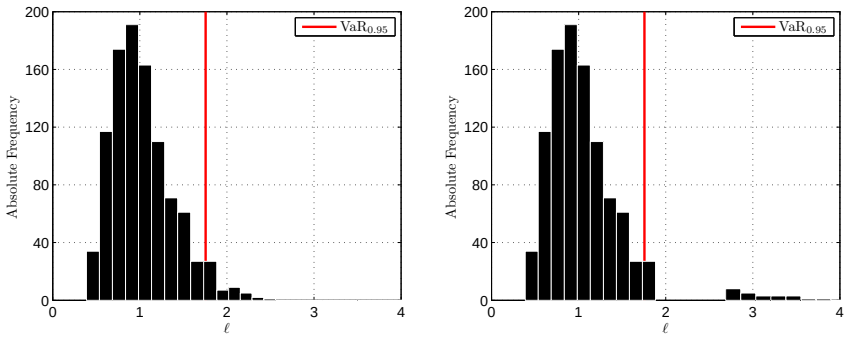


Figure 2.20: Value–at–Risk and extreme losses.

Another drawback of VaR is its non–coherence. Artzner et al. (1999) set up a set of requirements for a risk measure, $\delta(L)$. Those measures meeting all criteria are called "coherent" by Artzner et al. (1999). One of the properties a meaningful risk measure should have is the subadditivity property

$$\delta\left(\sum_{i=1}^{n} L_i\right) \leq \sum_{i=1}^{n} \delta_i \,. \tag{2.106}$$

A joint modeling of losses from distinct categories such as risk types or geographical locations, should thus not lead to an increase in risk as compared to summing up the single risk estimates. This property is, in general, not fulfilled by VaR; this drawback

will be discussed in more detail below.

A second risk measure, which has gained popularity recently, is given by the Expected Shortfall (ES)

$$\mathrm{ES}_\alpha = \mathrm{E}[L|L \geq \mathrm{VaR}_\alpha] \ . \tag{2.107}$$

This risk measure focuses on the tail of the loss distribution beyond VaR_α and, therefore, contains more information than the latter. Figure 2.21 illustrates this advantage of ES. Additionally, ES is a coherent risk measure in the sense of Artzner et al. (1999): It fulfills the subadditivity property (2.106) independently of the loss distribution's nature.

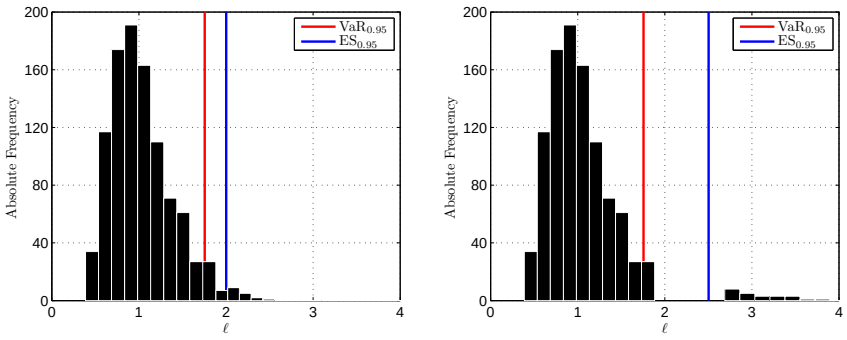


Figure 2.21: Value–at–Risk, Expected Shortfall and extreme losses.

Before the rise of VaR during the 90ies, variance was a very popular risk measure. For example, the pathbreaking approach to portfolio theory of Markowitz (1952) relies on variance. However, it is a symmetric measure covering the entire range of the distribution which may be a clear disadvantage in cases where tails are the focus of analyses and/or loss distributions are asymmetric. Furthermore, the existence of variances are not guaranteed for financial data due to the typical fat–tailedness of return (or loss) distributions. There exist other risk measures, such as Lower and Upper Partial Moments (Fishburn, 1977; Bawa, 1975, 1978) or Maximal Drawdown (Chekhlov et al., 2005). As these are less common than VaR and ES in the risk–management framework, we only consider the latter two when assessing effects of dependencies on risk capital.

2.4.2 Risk Aggregation and Dependence

Very often, there is a need to consider aggregate risk. For example, a company might desire to obtain one risk–capital estimate for all of market, credit, operational and other

risks, or there might be the need to aggregate over several separate models for one risk type, for example, at different geographical locations. In these cases, some sort of risk aggregation has to take place. In general, two different approaches may be used to obtain a risk–capital estimate $\delta(L)$ for aggregate losses $L = \sum_{i=1}^{n} L_i$ over i partitions. Firstly, risk might be modeled separately within each of the $i = 1, \ldots, n$ subcategories; in this case, n risk–capital estimates $\delta(L_i)$ will be the result. In a second step, these quantities have to be aggregated, which typically boils down to a summing–up of the single estimates. Secondly, the subcategories might be modeled jointly, so that one considers only one loss distribution, $F(L)$. In this case, risk–capital follows directly from the aggregate loss distribution.

There is no general answer to the question of which approach is preferable, or which one will lead to higher risk–capital estimates. Subadditivity, as required by Artzner et al. (1999) and introduced in Equation (2.106), implies that the first ("bottom–up") approach constitutes an upper bound for the results from the second ("bird's eye") approach: In the bivariate case, subadditivity implies that

$$\delta\left(L_1 + L_2\right) \leq \delta(L_1) + \delta(L_2)\ . \tag{2.108}$$

As pointed out by Artzner et al. (1999), this is intuitively appealing in the sense that *"...a merger does not create extra risk"*. This might be debatable in a context where a merger leads to higher systematic risks through concentration, for example, towards industry effects, weather conditions or the political environment. However, in a diversification sense similar to classical portfolio analysis, requirement (2.108) appears realistic.

While property (2.108) is always fulfilled for ES, this is not in general true for VaR, but only for special cases. If the underlying distribution of losses, L, is elliptical, VaR is subadditive for confidence levels $\alpha \in [0.5, 1)$ (see Theorem 6.8 of McNeil et al., 2005). Moreover, it can be shown that for comonotonic risks,

$$\mathrm{VaR}_{\alpha}^{\mathrm{co}}(L_i + L_j) \quad = \quad \mathrm{VaR}_{\alpha}(L_i) + \mathrm{VaR}_{\alpha}(L_j) \tag{2.109}$$

holds, so that VaR is additive in this special case. In the popular elliptical—and, thus, also in the Gaussian—world, comonotonicity translates into perfect positive correlation due to linearity. Therefore, once again, elliptical distributions lead to intuitively appealing and analytically tractable results.[23] This intuition may, however, be very

[23] One can therefore argue that VaR is a very useful risk measure for elliptical distributions. This shows that VaR *per se* is not a useless "model", as was often stated in the context of the recent financial crisis.

misleading outside the world of elliptical distributions. In fact, for non–elliptical distributions, it may happen that

$$\mathrm{VaR}_\alpha(L_1 + L_2) > \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2) \,, \tag{2.110}$$

the reason being the well–known lack of subadditivity of the VaR measure. This has been pointed out, for example, by Embrechts et al. (2002). As shown by Embrechts et al. (2009), a superadditive VaR may result, for example, from fat–tailed marginal distributions or restrictions of the joint distribution to the positive quadrant.

2.4.3 Bounds on Aggregate Risk

As comonotonicity does not always represent an upper bound on the VaR of the distribution of aggregate losses, one may ask whether such bounds exist at all. If they do, worst–case–scenarios for VaR can be constructed independently of the nature of the underlying distributions. This is a well–known issue—also known as the Fréchet Problem—attributed to Kolmogorov and first treated by Makarov (1981). It was taken up by Frank et al. (1987) and recently by Embrechts et al. (2003) and Embrechts and Puccetti (2010).

In a generalized context, one typically considers random variables $X_1, \ldots, X_d$ with marginal distribution functions $F_1, \ldots, F_d$ and a function $\Psi : \mathbb{R}^d \to \mathbb{R}$ of these d different types of risk. In our application, $\Psi(x_1, \ldots, x_d) = \sum_{i_1}^{d} x_i$, and $d = 2$. Furthermore, we denote by $\Psi_{x_1,\ldots,x_{d-1}}$ the function with variables x_1 to x_{d-1} held fixed, and by $\Psi^\wedge(u)$ the generalized right–continuous inverse of $\Psi(\cdot)$.

In order to state these bounds, three functions need to be defined. The supremal convolution of the d marginal variables is given by

$$\tau_{C,\Psi}(F_1, \ldots, F_d)(s) = \sup_{x_1,\ldots,x_{d-1}\in\mathbb{R}} C\big(F_1(x_1), \ldots, F_{d-1}(x_{d-1}), F_d(\Psi^\wedge_{x_1,\ldots,x_{d-1}}(s))\big) \,. \tag{2.111}$$

Correspondingly, the infimal convolution is obtained from

$$\rho_{C,\Psi}(F_1, \ldots, F_d)(s) = \inf_{x_1,\ldots,x_{d-1}\in\mathbb{R}} C\big(F_1(x_1), \ldots, F_{d-1}(x_{d-1}), F_d(\Psi^\wedge_{x_1,\ldots,x_{d-1}}(s))\big) \,, \tag{2.112}$$

and finally, the σ–convolution is defined as

$$\sigma_{C,\Psi}(F_1, \ldots, F_d)(s) = \int_{\{\Psi\le s\}} \mathrm{d}C\left(F_1(x_1), \ldots, F_d(x_d)\right) \,. \tag{2.113}$$

If $X_1, \ldots, X_d$ have marginal distribution functions $F_1, \ldots, F_d$ and copula C, the σ–convolution equals the distribution function of $\Psi(X_1, \ldots, X_d)$, $F_{\Psi(X_1,\ldots,X_d)}$.

The crucial point in deriving universal bounds on VaR are the Fréchet–Höffding bounds given in Equation (2.43). As these always hold, they imply bounds on VaR which are always satisfied. If the dependence structure as given by the copula can further be restricted, the bounds on VaR will, correspondingly, become narrower. In general, one has to assume that the copula C of $X_1, \ldots, X_n$ satisfies $C \geq C_0$ and $C^d \leq C_1^d$, where C^d is the dual of the copula C, i.e., $C^d(u_1, u_2) = u_1 + u_2 - C(u_1, u_2)$.

Under this assumption, the following inequality holds:

$$\tau_{C,\Psi}(F_1, \ldots, F_d)(s) \leq \sigma_{C,\Psi}(F_1, \ldots, F_d)(s) \leq \rho_{C,\Psi}(F_1, \ldots, F_d)(s) \; . \tag{2.114}$$

The supremal and infimal convolutions thus represent lower and upper bounds to the distribution of $\Psi(X_1, \ldots, X_d)$. Fixing a confidence level α, these bounds can be expressed in terms of quantiles, i.e.,

$$\rho_{C,\Psi}(F_1, \ldots, F_d)^{-1}(s) \leq \mathrm{VaR}_\alpha(\Psi(X_1, \ldots, X_d)) \leq \tau_{C,\Psi}(F_1, \ldots, F_d)^{-1}(s) \; . \tag{2.115}$$

The computational problem connected to Equation (2.116) consist of the unboundedness of the set $\mathbb{R}^{n-1}$ over which infima and suprema have to be taken.[24] Making use of the duality principle of Frank and Schweizer (1979), the bounds (2.116) can be stated in an alternative way which allows their computation. That is, denoting $\rho_{C,\Psi}(F_1, \ldots, F_d)^{-1}(s)$ by $F^{-1}_{\max}(\alpha)$ and $\tau_{C,\Psi}(F_1, \ldots, F_d)^{-1}(s)$ by $F^{-1}_{\min}(\alpha)$, one can write

$$F^{-1}_{\min}(\alpha) = \inf_{C_0(u_1,\ldots,u_d)=\alpha} \Psi\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) \; , \tag{2.116}$$

$$F^{-1}_{\max}(\alpha) = \sup_{C_1^d(u_1,\ldots,u_d)=\alpha} \Psi\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_d)\right) \; . \tag{2.117}$$

For the two–dimensional case, to which we are restricting attention, $C \geq C_0$ implies that $C^d \leq C_0^d$, so that we can take $C_0 = C_1$. The more restrictive the assumptions on the copula bounds, the tighter the bounds on VaR will be.

The resulting bounds will be used in Chapter 5 in order to derive worst–case scenarios for VaR.

[24] The computational aspects of the Fréchet problem are treated in detail by Williamson and Downs (1990).

Chapter 3

Quantifying Market Risk

3.1 The Market Risk Challenge

3.1.1 Univariate Stylized Facts

Financial time series have certain patterns in common. There is a long list of such "stylized facts" which apply to almost any price or return series. As we are interested in modeling the dependence structure of returns there is, at first sight, no reason to consider these peculiarities of the marginal distributions. However, there are at least two properties which have a direct implication for dependency modeling.

Firstly, one typically detects non–normality of financial return distributions. That is, they assign a larger probability to extremes than the normal distribution (fat–tailedness), and their tails decay according to a power law.

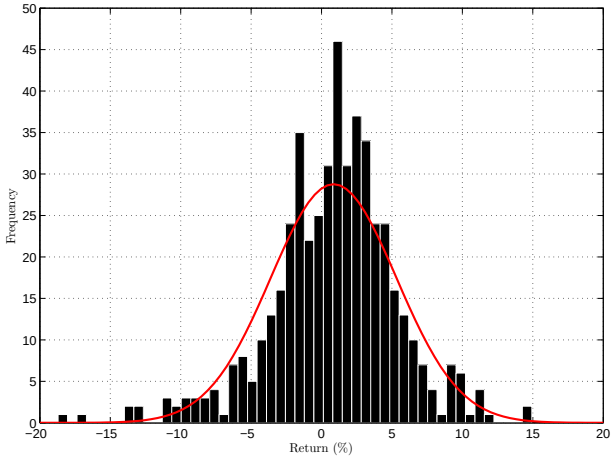


Figure 3.1: Histogram of Monthly Returns on the MSCI World Index, 1970–2010.

Figure 3.1 shows a histogram of monthly returns on the MSCI World Index between February 1970 and October 2010; the normal distribution estimated from these data is superimposed as red line. We see that extreme return values such as -17% are not adequatly captured by the normal distribution. This fat–tailedness, jointly with the fact that the empirical distribution is more peaked in the center, is known as leptokurtosis. This stylized fact has important implications for risk management, as it implies that under the assumption of the normal distribution the risk of losses may be severely underestimated. Apart from this univariate argument, the deviation from the normal also casts doubt on the multivariate normal as an appropriate distribution for joint return movements. As we will see later on, this problem of underestimating risk by using the normal distribution also transfers to the multivariate framework.
Besides their fat–tailedness, (univariate) financial returns tend to be asymmetric (Aït-Sahalia and Brandt, 2001)—a property that the normal distribution cannot capture either.
A second typical (univariate) property of financial return series refers to their dynamic behavior and is known as volatility clustering. Figure 3.2, containing the time series of

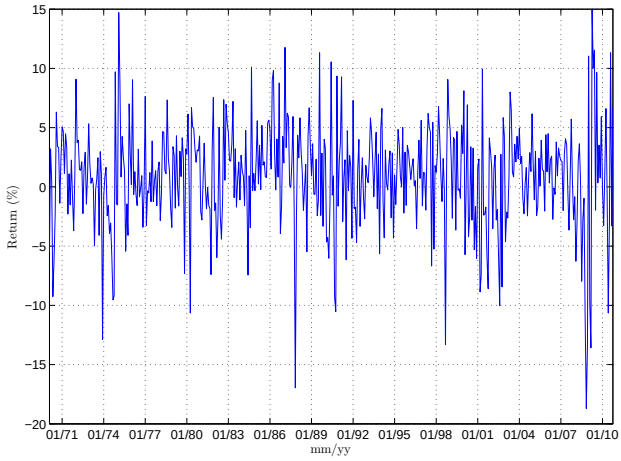


Figure 3.2: Time Series of Monthly Returns on the MSCI World Index, 1970–2010.

returns on the MSCI World Index, illustrates that periods of large (positive or negative) returns tend to be followed by periods with large (negative or positive) returns—the same holding for periods with only small movements in prices. This serial dependency in volatilities, which is known as GARCH effect, is best seen when comparing the relatively tranquil period between 2004 and 2007 to the extreme movements observed afterwards during the financial crisis.

There exists a vast literature on models extending the basic GARCH model of Bollerslev (1986); for a comparison of 330 different volatility models, see Hansen and Lunde

(2005). An overview and in–depth discussion of stylized facts as well as modeling under non–Gaussian distributions can be found in Jondeau et al. (2006).

3.1.2 Correlation Breakdown

Diversification in the classical mean–variance model of Markowitz (1952) is illustrated in Figure 3.3 for a simple, bivariate setting. Considering two assets characterized by expected returns μ_i and standard deviations σ_i, $i = 1, 2$, the expected return of a portfolio of these two assets, each one with weight x_i, is given by

$$\mu_p = x_1\mu_1 + x_2\mu_2 \ . \tag{3.1}$$

The portfolio's standard deviation is accordingly given by

$$\sigma_p = \sqrt{x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_ix_2\sigma_1\sigma_2\rho} \ , \tag{3.2}$$

where ρ denotes the correlation between the two assets' returns. Figure 3.3 shows the resulting combinations of portfolio expected return and risk (the latter being measured by standard deviation) for different values of ρ.

Obviously, for a perfect positive correlation, the possible portfolios represent linear combinations in the (σ_p, μ_p)–space. In other words, no diversification can be achieved. For lower levels of correlation, we see that the diversification benefit increases; that is, for a given level of expected portfolio return, the portfolio standard deviation decreases. This very basic idea suggests an investment in assets characterized by low correlations in order to reduce risk as much as possible. Such a low correlation may, for example, be due to different geographical locations (Levy and Sarnat, 1970; Solnik, 1974) or different asset classes such as gold and stocks (Chua et al., 1990).

However, an adoption of this simple approach may bear the risk of overestimation of the diversification effect. Firstly, it is well–known that return series of financial data tend to exhibit fat tails. These cannot be captured by the mean–variance approach, which relies exclusively on the first two moments and thus on the normal distribution. From a multivariate viewpoint, nonlinear dependence patterns such as tail dependence may render linear correlation a dangerous measure. Secondly, even if returns were normally distributed, one could suspect correlations not to remain constant over time. Intuitively, we expect that through times of crises, when common economic factors and herding drive returns to a large extent, correlations among the components of a portfolio increase. In fact, this so–called correlation breakdown or contagion effect is typically found in the data and has become a common idea both among practitioners

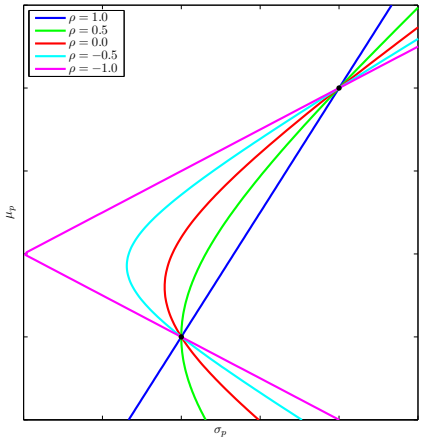


Figure 3.3: Diversification in a Two–Asset Portfolio.

and academics.[1]

There exists a strand of literature which deals with the question of how correlations among assets change in turbulent times. Earlier work divided the sample into distinct periods and tested for equality of the correlation matrices in these subsamples. For example, analyzing international diversification, Kaplanis (1988) finds evidence for stability of the correlation matrix over (predefined) subperiods; see also Ratner (1992) and the references therein. On the other hand, Koch and Koch (1991) and King et al. (1994), among others, find evidence for an increase in correlation among national markets over time, which is typically argued to be caused by increasing globalization and integration of markets.[2]

In order to identify correlation breakdown, it is necessary to distinguish between unconditional and conditional correlation. The correlation breakdown—or contagion—hypothesis refers to changes in linkages among markets; it thus concerns unconditional correlations. Assessing such changes is inherently difficult, as the consideration of a (sub)sample from the underlying distribution always induces a sort of conditioning on market observables. In general, changes in correlation across different samples can be due to two effects. Firstly, it might be that expected returns and variances change over time, while the underlying dependency structure remains constant; in this case, one would observe a spurious relationship between correlation and volatility. Secondly, the

[1] At the time of writing, there is no common definition of correlation breakdown; see, for example, Dornbusch et al. (2001) for a review.

[2] For further—partially conflicting—studies on the question of the stability of correlations among international indices, see, for example, Panton et al. (1976), Hilliard (1979), Watson (1980) and Maldonado and Saunders (1981). Recently, Ronn et al. (2009) have found evidence for correlation breakdown on stock markets.

dependence structure itself may vary. In both cases, one observes changes in conditional correlations, while only the latter reasoning corresponds to differences in unconditional correlations and thus to the notion of correlation breakdown.

In order to disentangle these effects, one could try and capture volatility dynamics; remaining changes in conditional correlation would then hint at the presence of correlation breakdown. This is done by Longin and Solnik (1995), who find evidence for changing conditional correlations beyond the GARCH model.[3] Using monthly excess returns for seven countries between 1960 and 1990, they find that correlations tend to increase over time, are higher in cases of large shocks, and are linked to interest rates. Karolyi and Stulz (1996) confirm the effect of large shocks, but do not find sufficient evidence for the role of macroeconomic announcements, industry effects or interest rate shocks.

However, when using conditional correlations as indicators for correlation breakdown, care has to be taken as soon as variances are affected by conditioning. In fact, as first pointed out by Boyer et al. (1999), considering subsamples obtained via conditioning may lead to changing correlations even if the sample is generated from a common distribution with constant parameters. In order to illustrate this, we give an example based on the ideas of Boyer et al. (1999).

Example 3.1. *We draw 10,000 data points from a bivariate normal distribution with (unconditional) correlation coefficients* $\rho = 0.1, 0.3, 0.5, 0.8$. *The conditional correlation between* X_1 *and* X_2 *is then estimated from subsamples: We select those data points for which* x_1 *is either below the* $\alpha/2$*–quantile or above the* $1 - \alpha/2$ *quantile, using* $\alpha = (0.01, 0.05, 0.1, 0.15, 0.2, \ldots, 0.5$. *For example, taking* $\alpha = 0.1$ *means conditioning on* x_1 *being either below the 5% quantile or above the 95% quantile.*

The resulting conditional correlations are shown in Figure 3.4 for different values of the unconditional correlation. For the 50% *quantile, the conditional correlation boils down to the unconditional one, as* X_1 *can only be below or above the median. However, decreasing* α*—that is, diminishing the conditioning tail area of* X_1*—leads to an increase in conditional correlation.*

The relationship between unconditional and conditional correlation depends on unconditional and conditional variances. As is shown by Boyer et al. (1999), for two normally distributed random variables with unconditional correlation ρ the correlation

[3] Ang and Bekaert (2002) and Ang and Chen (2002) confirm that (symmetric as well as asymmetric) multivariate GARCH models cannot fully capture changes in conditional correlations.

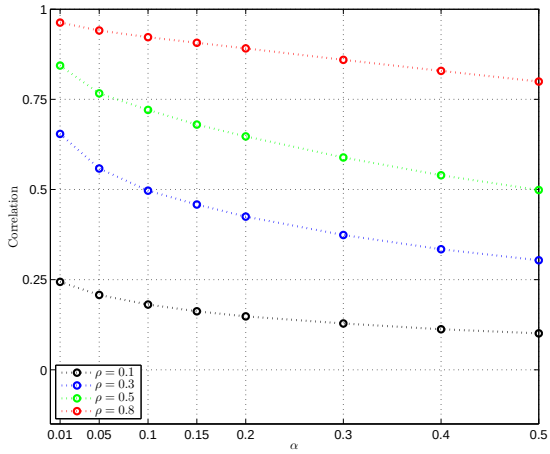


Figure 3.4: Conditional Correlations for the Multivariate Normal: Two–Sided Conditioning

conditional on occurrence of event $\mathcal{A}$ is given by

$$\rho_{\mathcal{A}} = \rho \left(\rho^2 + (1 - \rho^2) \frac{\text{Var}[X_1]}{\text{Var}[X_1|\mathcal{A}]} \right)^{-1/2} . \tag{3.3}$$

For the two–sided conditioning illustrated in Figure 3.4,

$$\text{Var}[X|\mathcal{A}] \geq \text{Var}[X] \ , \tag{3.4}$$

simply because observations in the tails are spread more widely. The difference between unconditional and conditional variance is the higher, the smaller α; accordingly, equality corresponds to the case $\alpha = 0.5$. This relationship implies, via Equation (3.3), that $\rho_{\mathcal{A}} \geq \rho$, which is exactly what we observe in Figure 3.4. In general, the conditional correlation between X_1 and X_2 is larger (smaller) than the unconditional correlation if the conditional variance of X_1 given $\mathcal{A}$ is larger (smaller) than the unconditional variance. Boyer et al. (1999) also consider one–sided tail conditioning as well as a direct conditioning on volatilities; obviously, Equation (3.3) also holds in these cases, so that conditional correlations differ from their unconditional counterparts.

Example 3.2. *To illustrate the effect of conditioning via variances, we draw 100,000 bivariate data points* (X_i, Y_i), $i = 1, \ldots, 100{,}000$, *from a multivariate normal distribution with unconditional correlation* $\rho = 0.4$ *and build groups of 100 observations each. For each of these groups, we calculate the sample variance. Conditional correlations are evaluated from a subsample containing those groups for which the variance of* X_1 *exceeds a certain threshold. Figure 3.5 shows the resulting conditional correlations for different values of thresholds.*

We see that conditional correlations increase as the minimum variance increases, although the underlying distribution is identical with a constant correlation.

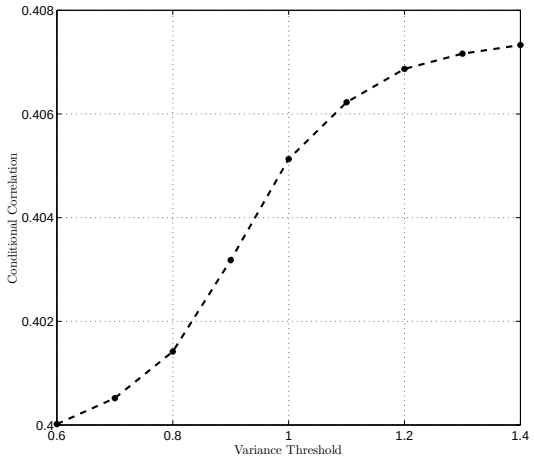


Figure 3.5: Conditional Correlations for the Multivariate Normal: Conditioning via Variances

The mere fact of observing changing correlations over subsamples obtained via conditioning does, therefore, not have to be caused by correlation breakdown. Rather, the changing correlation may be a result of the selection bias induced by consideration of subsamples. Loretan and English (2000) generalize the result of Boyer et al. (1999) to the case where one random variable is an affine function of the other. Forbes and Rigobon (2002) suggest a bias correction adjusting for the effect of heteroscedasticity; analyzing stock market returns in Hong Kong and the Philippines during 1997, they find evidence for correlation breakdown in conditional correlations which basically vanishes after bias correction. This finding questions the existence of a correlation breakdown effect and leads the authors to conclude that there is "no contagion, only interdependence", the latter referring to cross–market linkages which may be strong, but are constant over time.

Although changes in the level of correlation among subsamples *per se* can not be used to infer a correlation breakdown, the form of the change can hint at the appropriateness of constant unconditional correlation and the underlying normality assumption. Conditioning on exceedance of VaR, Campbell et al. (2002) follow this idea by comparing empirical and theoretical quantile–based correlations. Longin and Solnik (2001) introduce the notion of exceedance correlation. For two random variables X_1, X_2, this quantity is given by

$$\rho(\theta) = \begin{cases} \text{Corr}\left[x_1, x_2 | x_1 < \theta, x_2 < \theta\right] & \text{for} \quad \theta \leq 0 \\ \text{Corr}\left[x_1, x_2 | x_1 > \theta, x_2 > \theta\right] & \text{for} \quad \theta \geq 0 \end{cases} . \tag{3.5}$$

In other words, correlation is calculated conditional on the joint exceedance of a threshold θ. This is illustrated in Figure 3.6, where the shaded region is the range of ob-

servations used for calculating $\rho(\theta)$. In contrast to Example 3.1, conditioning now

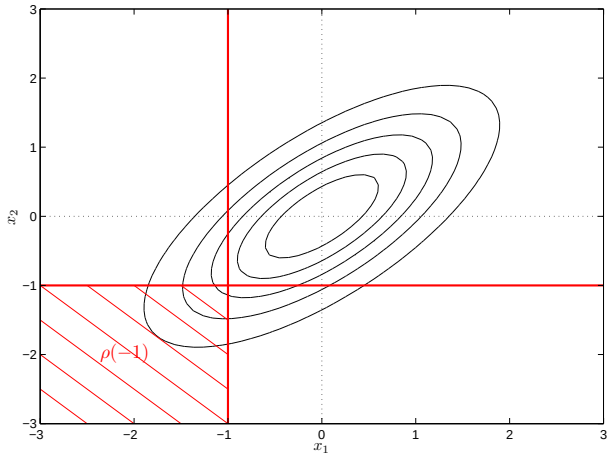


Figure 3.6: Exceedance Correlation: Illustration

affects both random variables; furthermore, it is a one–sided concept in that the signs of returns are relevant. The latter point makes exceedance correlations a useful tool for studying asymmetries in dependence; that is, it allows to compare the dependency structures in the tails of the distribution. However, observing differences in $\rho(\theta)$ for different values of θ does, again, not imply a change in the underlying distributional properties, as can be seen from Figure 3.1. It shows the behavior of exceedance correlation for the multivariate normal distribution with different values of ρ. We see

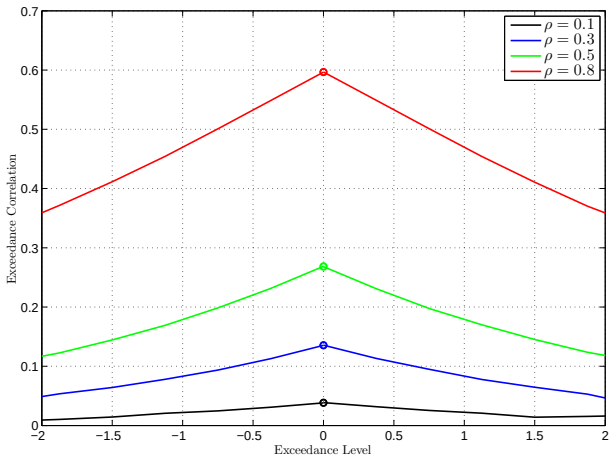


Figure 3.7: Exceedance Correlations for the Multivariate Normal

that, under the normal distribution, exceedance correlations decrease, the further one moves into the tail. This result is at first counter-intuitive, even more when considering the increasing conditional correlations of Example 3.1: In fact, we would expect

exceedance correlation to increase with the size of return. However, exceedance correlation is closely linked to the concept of tail dependence, and the multivariate normal distribution is asymptotically independent for $-1 < \rho < 1$. Therefore, exceedance correlations must decrease.

Rather than concluding a correlation breakdown effect from the observation of changing correlation estimates, one can therefore compare exceedance correlation behavior under the multivariate normal distribution to the empirical one and thereby assess whether the two are in line. This has been done, for example, by Longin and Solnik (2001), who find that (exceedance) correlations among international equity index returns increase in bear, but decrease in bull markets, so that the multivariate normal distribution is appropriate for positive, but not for negative price movements. This asymmetry, in turn, implies that volatility alone cannot be the driving force of correlation changes. Furthermore, Corsetti et al. (2005) argue that the conclusion of absence of contagion drawn by Boyer et al. (1999) and Forbes and Rigobon (2002) is obtained from the restriction of the variances' effects on returns to the common factor, neglecting country–specific variance effects and leading to biased estimation of conditional correlation.

The failure of the multivariate normal distribution to capture asymmetric dependence structures has been addressed in different ways. For example, multivariate GARCH models allowing for asymmetries (see, for example, Ang and Chen, 2002) or regime–switching models (Ang and Bekaert, 2002) have been applied in order to explain correlation breakdown. However, as is shown by Garcia and Tsafack (2009), multivariate GARCH and regime–switching models with Gaussian innovations cannot reproduce *asymptotic* asymmetries in dependencies; the asymmetries produced vanish asymptotically. We therefore take a different road in the following, questioning the appropriateness of correlation as dependence measure in the context of joint bear market movements. Instead, we aim at modeling tail dependence via copulas.[4] That is, we aim at estimating dependencies among extremes without letting our results be affected by the marginal distributions; furthermore, we want to be able to assess exceedances of extremes which may not be part of the underlying sample. This is only possible with an asymptotic concept such as tail dependence.

[4]This view has been approached, for example, by Patton (2004) or Hu (2006).

3.2 Correlation Breakdown Across Asset Classes

The correlation breakdown effect has been mostly studied in the context of international diversification. Early studies like those of Grubel (1968) and Agmon (1972) had—based on analyses of correlation coefficients among stock market indices—suggested such an investment strategy. Furthermore, crises such as the Mexican in 1994 or the "Asian Flu" starting in 1997 have been found to propagate quickly via international markets to other countries, which also catalyzed the analysis of changing dependence structures among national markets during turbulent times. Therefore, the conceptual problems of correlation breakdwon, their implications as well as empirical evidence have mostly been assessed across in the framework of international portfolio choice.

However, the diversification argument applies equally well when considering different asset classes. For example, one would suspect that in times of turbulences at stock markets, gold provides protection against large portfolio losses due to the correlation among stocks and gold. We, therefore, aim at assessing the dependency structure among three different indices with respect to correlation breakdown.

3.2.1 The Data

We consider monthly rates of return of three different indices between February 1970 and October 2010. The stock market is represented via the MSCI Total Return World Index, which measures the market performance in terms of prices and dividends, of about 1,500 stocks from 24 countries (excluding emerging markets). We include commodities via the S&P GSCI Commodity Total Return Index, which measures the returns accrued from investing in currently 24 fully-collateralized nearby commodity futures from the energy, industrial and precious metals, agriculture and livestock sectors. Gold is a third asset class promising substantial diversification benefits; its price per ounce is given by the Gold Bullion LBM U$/Troy Ounce Index.

Figure 3.8 shows the three indices over the time horizon considered in our analyses. Looking at their behavior, we detect several instances where opposite price movements might have provided protection against large portfolio losses. For example, commodities experienced a substantial decrease between January 1997 and April 1999, while stocks continued to increase—besides a small decline—until May 2000. Similarly, while large losses both in stocks and commodities can be observed during the recent financial crisis starting in 2008, the gold price exhibits only a minor decrease. These different asset classes thus suggest themselves to risk–averse investors for the purpose of diversification.

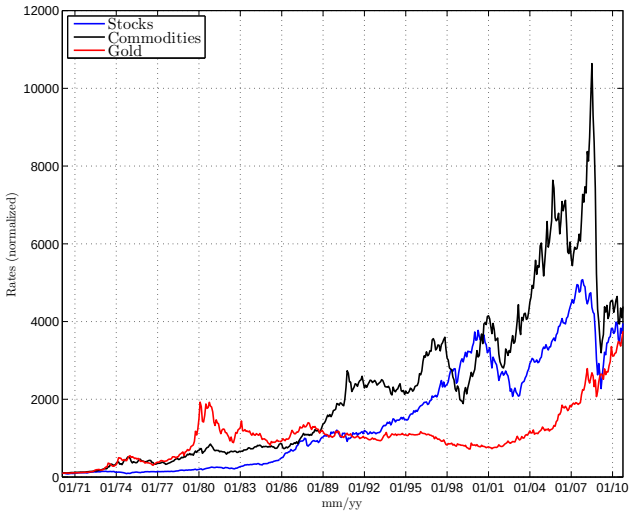


Figure 3.8: Normalized Rates of the Considered Asset Classes

3.2.2 Correlation

The most obvious attempt at modeling possible diversification benefits consists of estimating unconditional correlations. For the (raw) return series shown in Figure 3.9, the resulting linear correlation coefficient is shown in the first column of Table 3.1.

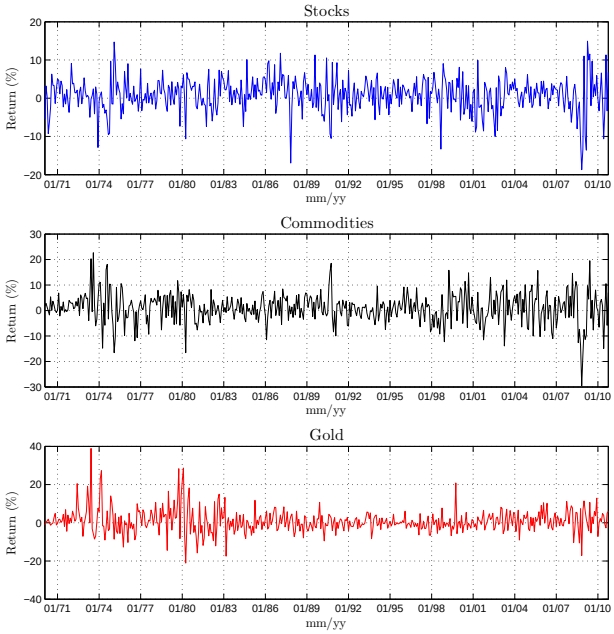


Figure 3.9: Returns of the Considered Asset Classes

We find the lowest correlation—and thus the highest potential for diversification—between stocks and gold. However, the correlation between stocks and commodities is only slightly higher. The strongest link is found between commodities and gold, which is not surprising, since gold is one of the commodities included in the S&P GSCI index. The second and third column of Table 3.1 contains estimates for Kendall's τ and Spearman's ρ_S. Rank correlation estimates differ in values, but not in the relative ordering with respect to the riskiness of asset class combinations.

Pair	ρ	τ	ρ_S
Stocks and Commodities	0.1722	0.0946	0.1360
Stocks and Gold	0.1101	0.0553	0.0834
Commodities and Gold	0.2725	0.1562	0.2274

Table 3.1: Unconditional Correlation Estimates

However, correlations among asset classes may vary substantially, depending on the subsample chosen. As an illustration, Figure 3.10 shows linear correlations obtained from a one–year rolling window; that is, the first value (01/71) is calculated from returns between February 1970 and January 1971, the last one (10/10) is based on returns between November 2009 and October 2010. For all three asset class combinations, we observe large variations in correlation.

One finds periods in which correlation is clearly negative, as is the case for stocks and commodities in January 1976. At the same time, and more dangerously, there are periods which are characterized by large positive correlations. For example, between January 1980 and September 1981—thus, in the aftermath of the second oil crisis—all asset class combinations show a high positive correlation. Looking at Figure 3.9, we confirm that this period is one of high variability in returns. Similar effects are observed, for example, for the correlation between stocks and commodities after the bursting of the dotcom bubble in 2000, as well as during the recent financial crisis starting in the fourth quarter of 2008. In other words, we find evidence for correlation breakdown: During the turbulent times when diversification is most needed, it fails to protect against large portfolio losses. Using an unconditional correlation estimate as those of Table 3.1 may therefore be too optimistic.

Another hint at the appropriateness of a constant correlation capturing the entire dependency structured is given by the exceedance correlations. These are shown in Figure

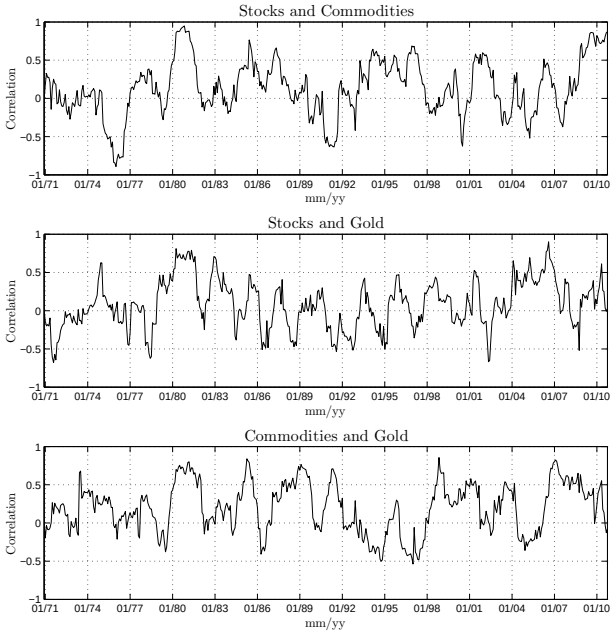


Figure 3.10: Linear Correlations, 12–Month Rolling Window.

3.11, where the empirical exceedance correlations are compared to those implied by the multivariate normal distribution. Here and in the following, we consider GARCH(1,1) filtered returns in order to circumvent effects of volatility clustering on the analyses of dependence structures. Therefore, the range of returns—and, thus, exceedance levels θ in Figure 3.11— is smaller than in the unfiltered case. As soon as there are less than ten joint observations below θ (or above it, for $\theta > 0$), we stop decreasing (or increasing) θ in order to avoid exceedance correlation estimates to be driven by single observations.

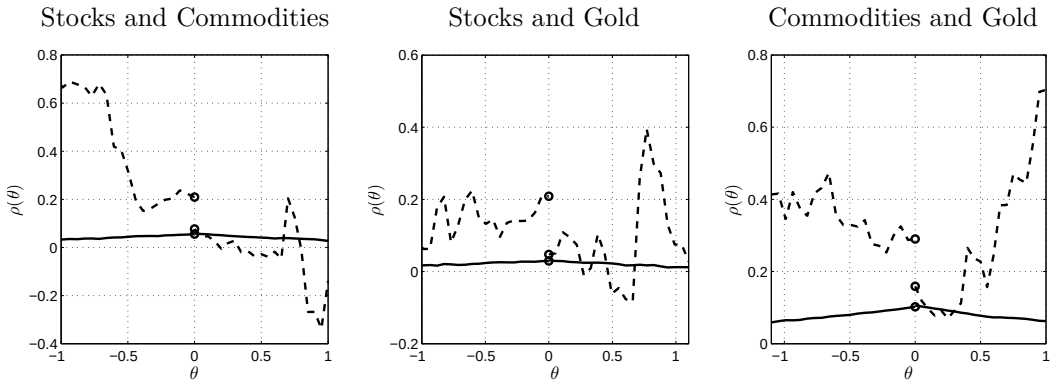


Figure 3.11: Empirical (Dashed) and Multivariate–Normal (Solid) Exceedance Correlations.

For all three asset class combinations, we find asymmetric exceedance correlation structures. That is, for negative θ—adverse price movements—conditional correlations tend to be higher than for positive returns. For example, given that the (GARCH–filtered) returns of stocks and commodities are both below 0.5%, their correlation is higher than 0.4. This value, when compared to an unconditional correlation coefficient of $\rho = 0.1722$ as reported in Table 3.1, illustrates once more the possible dangers of neglecting asymmetric correlation structures. However, we do not find such clear–cut relations as earlier studies did. For example, Longin and Solnik (2001) find exceedance correlations to increase in bear, but not in bull market. For our data, exceedance correlations for $\theta < 0$ increase only in two of three cases. Furthermore, we cannot find clear directions of changes for bull markets ($\theta > 0$). The pair coming closest to the exceedance correlations observed by Longin and Solnik (2001) consists of stocks and commodities.

The solid lines in Figure 3.11 are the exceedance correlations implied by a multivariate normal distribution fitted to the data. The deficits of the multivariate normal become apparent in two aspects: Firstly, the multivariate normal is not able to capture the asymmetries in correlation structures; secondly, and related to this, the level of exceedance correlation is underestimated by the multivariate normal. In fact, the implied exceedance correlations are at best as high as the empirical ones for $\theta > 0$, but do not come close to those for negative returns.

We conclude that going from an unconditional to a conditional correlation cannot save the assumption of multivariate normality. For the multivariate normal, (theoretical) exceedance correlations are not constant, so that the mere fact of observing changes in correlation does not imply the rejection of this distribution; rather, the structure of these changes for returns is so different from that of the multivariate normal that we have to seriously doubt the appropriateness of the latter distribution.

3.2.3 Copula Fitting

We fit several copulas to the filtered return series. In order to do so, we use the empirical distribution functions of the margins in order to avoid effects of misspecified margins on the parameter estimates. Table 3.2 shows the results. The first column contains the estimates; for example, the correlation coefficient of the Gaussian copula is equal to $\widehat{\rho} = 0.1306$ for stocks and commodities. Comparing the correlation parameter estimates of the Gaussian copula to the Pearson correlation coefficients of Table 3.1, we see that the relative ordering of asset combinations is the same. However, the values obtained from the Gaussian copula are lower than their counterparts, which can—at

least partially—explained by the fact that Table 3.1 is based on unfiltered returns and therefore may be influenced by GARCH effects.

	$\widehat{\Theta}$	$\widehat{\lambda_L}$	$\widehat{\lambda_U}$	AIC	BIC	p–value $\mathcal{A}_2$	p–value $\mathcal{A}_4$
	Stocks and Commodities						
Ga	0.1306	—	—	-6.0567	4.3280	0.7002	0.6875
t	8.0460	0.0442	0.0442	-11.5171	-1.1324	0.8993	0.5498
Gu	1.0739	—	0.0931	-3.3615	7.0232	0.1804	0.5357
GuS	1.0998	0.1219	—	-13.3075	-2.9228	0.9059	0.9305
Cl	0.1965	0.0294	—	-11.7214	-1.3367	0.7737	0.9994
ClS	0.0906	—	0.0005	-0.5572	9.8275	0.0177	0.2300
	Stocks and Gold						
Ga	0.0829	—	—	-1.2248	9.1599	0.6688	0.7843
t	35.9166	0.0000	0.0000	-1.4704	8.9144	0.8373	0.9860
Gu	1.0453	—	0.0592	-0.4839	9.9008	0.5652	0.9171
GuS	1.0502	0.0651	—	-2.2842	8.1005	0.6311	0.9923
Cl	0.0923	0.0005	—	-1.3470	9.0377	0.5204	0.9891
ClS	0.0730	—	0.0001	0.0783	10.4631	0.3738	0.7129
	Commodities and Gold						
Ga	0.2348	—	—	-24.6139	-14.2292	0.7892	0.9116
t	14.1987	0.0158	0.0158	-26.4730	-16.0883	0.9273	0.9431
Gu	1.1430	—	0.1661	-18.9034	-8.5187	0.0500	0.4639
GuS	1.1683	0.1901	—	-27.7348	-17.3501	0.8582	0.9917
Cl	0.2972	0.0971	—	-23.7156	-13.3308	0.3227	0.8207
ClS	0.2365	—	0.0533	-15.3145	-4.9298	0.0058	0.0222

Table 3.2: Copula Estimation Results

The second and third columns of Table 3.2 contain the lower and upper tail dependence coefficients implied by the parameter values of the first column. For example, the Student–t copula for stocks and gold has a high number of degrees of freedom ($\widehat{\nu} \approx 36$), meaning that it is in fact very close to the Gaussian copula and shows negligible tail dependence. Although the lower tail dependence coefficients thereby obtained are moderate, the AIC and BIC values in the next two columns favor those copulas

which capture lower tail dependence, thus hinting at the presence of it. For all three combinations, the Gumbel survival copula is the preferred one. The second– and third–best fit is obtained from the Student-t and Clayton copulas; only the ordering of these two changes. At this stage, we therefore conclude that the joint distributions of the returns show lower tail dependence, although its value is mostly moderate ($\widehat{\lambda_L} < 0.2$ in all cases). The highest tail dependence is found between stocks and commodities.

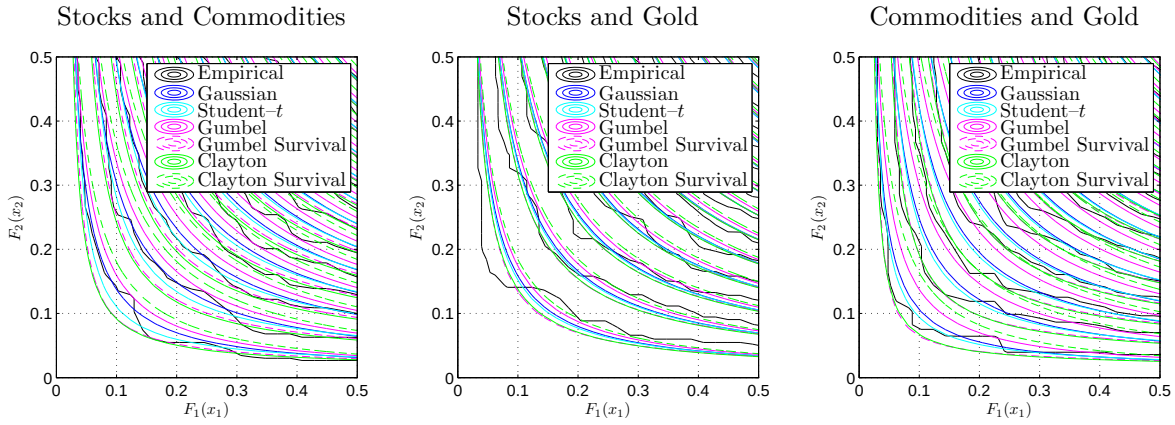


Figure 3.12: Contour Plots of Empirical and Fitted Copulas

Figure 3.12 shows level curves of the empirical and fitted copulas. As is to be expected, those copulas allowing for (lower) tail dependence (Gumbel survival and Clayton) lead to a higher probability of joint exceedance—in absolute terms—than those who don't (for example, the Gaussian) for some given $F_1(x_1)$, $F_2(x_2)$. Put differently, if we assume some given level of $\Pr[X_1 \leq x_1, X_2 \leq x_2] = C(F_1(x_1), F_2(x_2))$, the corresponding x_1 and x_2 will be lower for those copulas who model tail dependence.

The fact that copulas with asymmetric dependence structures and tail dependence fit better than the Gaussian copula becomes evident when revisiting exceedance correlations. Figure 3.13 shows the empirical exceedance correlations as seen in Figure 3.11 jointly with the exceedance correlations implied by the fitted copulas. The latter are obtained from simulated returns, using $B = 100{,}000$ replications and inversion of the empirical distribution functions. For readability reasons, we exclude those two copulas implying presence of upper and absence of lower tail dependence (Gumbel and Clayton survival), as their implied dependence structures are far from the empirical ones.[5]

[5] As the survival copula is the "mirror image" of the respective copula, the exceedance correlations of the two excluded candidate copulas correspond—in a qualitative sense—to the mirror image of their counterparts. That is, exceedance correlation is higher for positive returns.

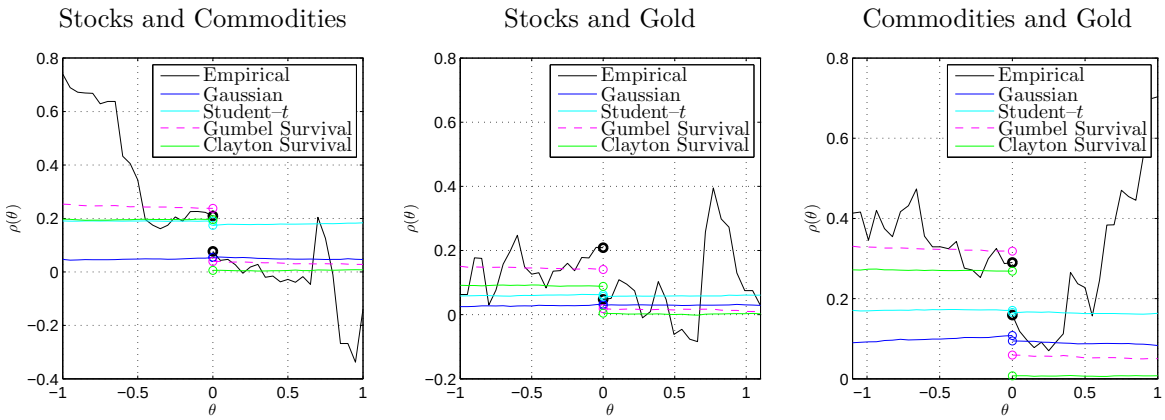


Figure 3.13: Empirical and Fitted Exceedance Correlations

As Figure 3.13 is obtained from simulations involving the empirical distribution, the exceedance correlations of the Gaussian copula are not perfectly symmetric. However, the deficits discussed with respect to the multivariate normal distribution apply here again: The asymmetric empirical structure cannot be captured, and correlations for negative values of θ are underestimated. The Student–t copula is able to overcome the latter point, as it captures tail dependence. Its implied exceedance correlations are therefore higher than those of the Gaussian copula, the difference between the two decreasing in the value of ν). That is, for stocks and commodities, where we obtained the lowest number of degrees of freedom, the difference between Gaussian and Student–t exceedance correlations is highest. However, the symmetry restriction is equally present under the Student–t copula: The fact that both positive and negative returns are required to be represented may lead to an overestimation of downside and an underestimation of upside diversification potentials. For example, for stocks and commodities, the Student–t implied exceedance correlations are mainly below the empirical ones for $\theta < 0$ and higher than empirical correlation for $\theta > 0$. This is overcome by the use of Archimedean copulas. As they allow the tail dependence coefficient to differ between the upper and lower tail, the exceedance correlations are closer to the empirical ones for $\theta < 0$. Overall, the results from Table 3.2 are confirmed: The Gumbel survival copula leads to the exceedance correlation structure which fits the data best. This is the same copula as used by Patton (2004) for capturing asymmetric dependencies.[6]

The copula approach, however, goes beyond what can be seen from exceedance correlations. In fact, just as linear correlation, exceedance correlations are affected by changes of the marginal distributions; this is not true for copulas and the implied tail

[6]Patton (2004) calls the Gumbel survival copula the "Rotated Gumbel".

dependence coefficients. Furthermore, the estimated tail dependence coefficients tell us something about asymptotic joint behavior of returns, which cannot be captured by exceedance correlations. At the same, we have shown that our copula approach can reproduce the exceedance correlation patterns typically observed.

3.2.4 Risk and Variations in Dependence Structures

In contrast to other risk types such as credit or operational risk, market risk is typically evaluated in terms of days. This is due to the fact that the instruments bearing market risks are typically constantly traded, their prices possibly showing a volatile behavior over relatively short time intervals.

We saw in Figure 3.10 that correlations may vary substantially over subsamples. We further found in Section 3.2.2 that the multivariate normal cannot explain this effect, but that copulas are able to capture changes and asymmetries in conditional correlations. One could therefore argue that failure of the multivariate normal (and, thus, correlation) is due to the fact that the underlying joint distribution involves a copula different from the Gaussian. However, it is not clear why this "true" copula should remain constant. Instead, its parameters and implied dependence structure may vary over time as well. Using the most popular correlation breakdown explanation in the context of copulas, we could suspect that lower tail dependence increases in turbulent times.[7]

In order to assess the variability of copula parameters and its implications on risk assessment, we consider subsamples of return series. Specifically, we use an rolling–window approach with window length $w = 72$ (corresponding to six years of data). This window length is chosen in order to balance a trade–off: If the subsample is too short, parameter estimates and implied dependence condiments are based on too few observations and very volatile. On the other hand, a very long subsample will dampen effects of extremes.

Starting with the subsample covering returns between Februay 1970 and January 1976, we estimate copula parameters and the multivariate normal distribution to the data in the subsample. We then calculate $\text{VaR}_{0.01}$ and compare it to the loss of the following period (starting in February 1976).

[7] This does, again, not exhaust the range of possible explanation. In fact, we could also suspect that correlations are regime–dependent, so that we have no longer one unconditional correlation, but rather several ones. See Haas et al. (2009) for a juxtaposition of (constant) copulas to a regime–switching model.

Stocks and Commodities. Figure 3.14 shows the variability of results. It shows the evolution of Pearson's correlation coefficient, the parameter estimate $\widehat{\alpha}$ of the Gumbel survival copula as well as the implied lower tail dependence coefficient over time. Turbulences which follow extreme return movements enter the estimation subsample only gradually and therefore reveal their full correlation–breakdown potential with a time lag.

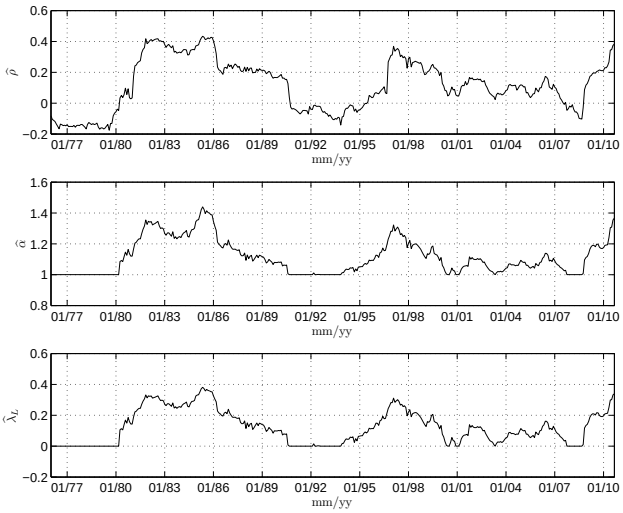


Figure 3.14: Time–Variation in Dependency, Stocks and Commodities.

This is, for example, seen in the rise in correlation only starting in 1980, when losses during the second oil crises enter the subsample. The maximum correlation is then reached in 1986, that is, one entire window length after. We observe a striking similarity between the three quantities. However, there is a difference in extremes: The increase in tail dependency may be sharper than that of correlation. This is seen, for example, in the periods 1984/195 and 1997/1998, where we observe a peak in tail dependence which is more pronounced than under correlation. This different behavior illustrates the danger of neglecting tail dependence: Although the dynamics of correlation and tail dependence are very similar, the increase in dependence in periods of extreme returns is underestimated by the multivariate normal distribution.

As these are the periods where diversification is needed the most, restricting attention to linear correlation may be misleading despite the similarity of changes in tail dependence and correlation. This is illustrated in Figure 3.15, which shows the $\text{VaR}_{0.01}$ obtained from both distributional assumptions together with returns for a portfolio equally divided between stocks and commodities.

The differences between the quantities is highest in those periods of extreme returns, for example, during the periods mentioned above. That is, around 1984/195 and 1997/1998, the $\text{VaR}_{0.01}$ under the Gumbel survival copula is clearly higher—in absolute value—than the one under the multivariate normal assumption, while the two estimated quantities are basically equal in other periods. The risk of underestimating potential losses in periods of turbulence is, therefore, lower when using a copula allowing for lower tail dependence.

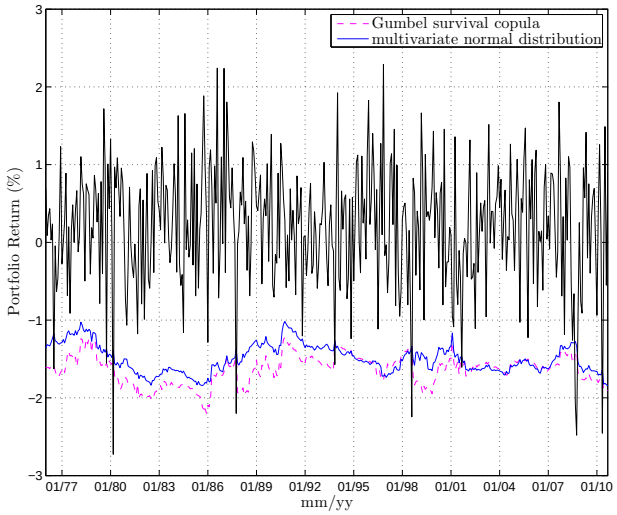


Figure 3.15: Portfolio Return, Stocks and Commodities.

Stocks and Gold. The patterns described above can similarly be found if we consider a portfolio equally divided between stocks and gold.

Large portfolio losses, such as observed for October 1987 or August 1998, lead to a much higher increase in $\text{VaR}_{0.01}$ under the Gumbel survival copula, hinting again at the relevance of lower tail dependence. For example, after the portfolio loss of October 1987, $\text{VaR}_{0.01}$ is around -2%; for the multivariate normal, it remains around -1.5%. However, we observe a new pattern for this portfolio: The order of riskiness is reversed between June 2002 and May 2008; during this period, the $\text{VaR}_{.01}$ under multivariate normality is lower than under the Gumbel survival copula.

Looking at the evolution of dependency measures in Figure 3.16, we see that this is caused by an increase in correlation which is (relatively) higher than that in lower tail dependence. Looking at Figure 3.17 reveals that this period is characterized by large positive returns and few portfolio losses; as soon as a large portfolio loss occurs (which

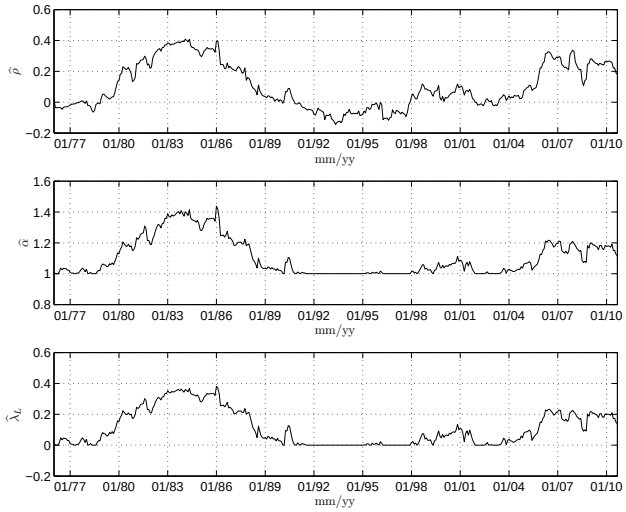


Figure 3.16: Time–Variation in Dependency, Stocks and Gold.

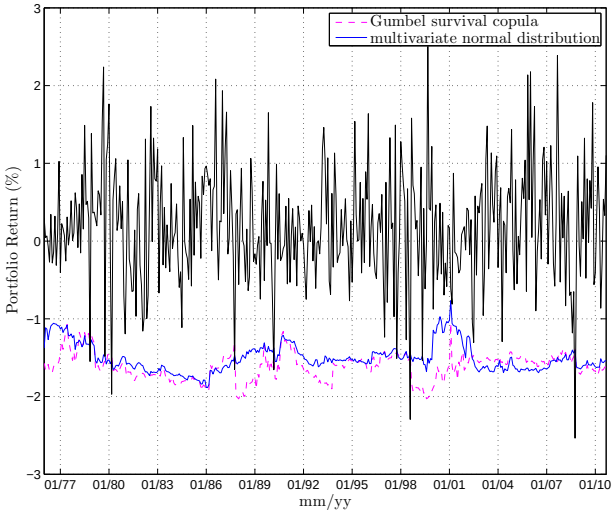


Figure 3.17: Portfolio Return, Stocks and Gold.

is the case in October 2008 with the Lehman bankruptcy), the Gumbel survival VaR is again more conservative than the multivariate normal one.

Commodities and Gold. The returns of an equally–weighted portfolio containing commodities and gold are shown in Figure 3.19.

Again, we find periods, such as between 1977 and 1980, where the $\text{VaR}_{.01}$ is higher—in absolute terms—under multivariate normality. Figure 3.18 shows that these months are characterized by positive dependence; for example, between 1977 and 1980, linear correlation varies between 0.4 and 0.2, and the upper tail dependence coefficient ranges from 0.15 to .25. Going back to Figure 3.9 and comparing the returns of commodi-

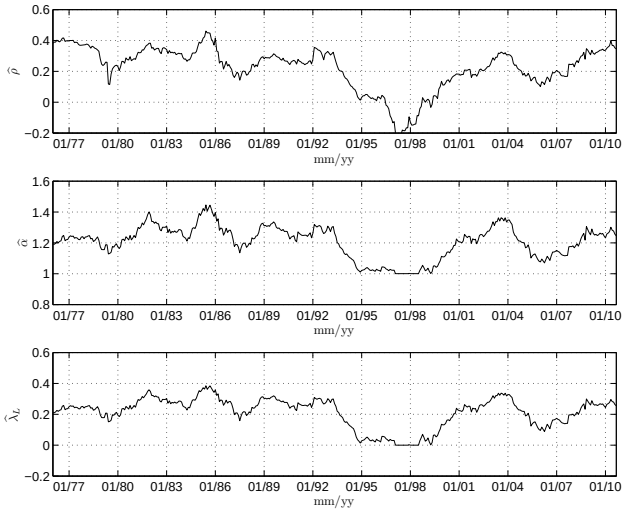


Figure 3.18: Time–Variation in Dependency, Commodities and Gold.

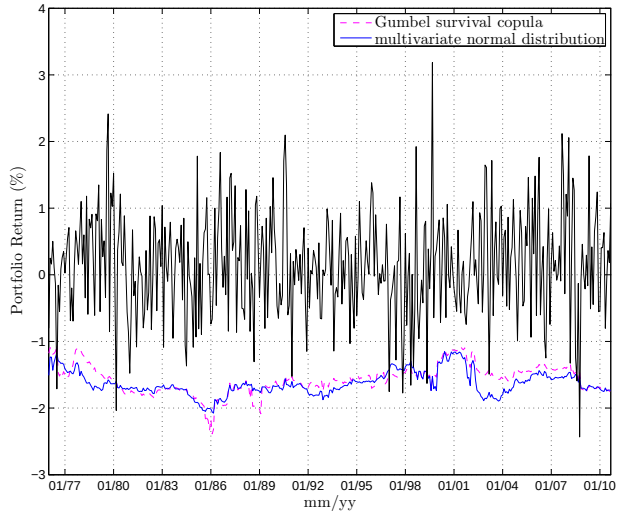


Figure 3.19: Portfolio Return, Commodities and Gold.

ties and gold until 1980, we see that the large movements are positive. For example, monthly returns were higher than 20% for commodities and up to almost 40% for gold around the first oil crisis. In other words, the correlation detected in the first subsamples comes from positive comovements of prices. As the normal distribution is symmetric, such large positive returns imply just the same (exceedance, or conditional) correlation as negative returns of the same size; the multivariate normal cannot distinguish between gains and losses when it comes to conditional correlations. On the contrary, the Gumbel survival copula implies lower tail dependence, but absence of upper tail dependence; therefore, the large positive movements as observed in the first subsamples have a smaller effect on VaR as negative movements of the same size would have. This is confirmed by the correlation and lower tail dependence coeffi-

cients around 1986—that is, in the periods where the second oil crisis its biggest effect. While the correlation coefficients around 1977 and 1986 are both around 0.4, the lower tail dependence coefficient is much higher in the oil crisis periods ($\widehat{\lambda}_L \approx 0.24$ versus $\widehat{\lambda}_L \approx 0.38$). Correspondingly, the $\text{VaR}_{.01}$ under the Gumbel survival copula is visibly lower than the multivariate normal one.

We conclude that the asymmetry in the Gumbel survival copula has an effect of down–weighting dependencies among positive returns, while capturing the lower tail dependence during crises. If the primary goal of dependency modeling is an adequate assessment of downside risk, it is thus preferable to the symmetric multivariate Gaussian.

The rolling–window approach to copula estimation provides us with an impression of changes in dependency structures over time. However, it cannot provide us with a prediction of crises, as these changes are not modeled explicitly. This can be achieved, for example, using regime–switching models where the copula differs between regimes (Balyeat and Muthuswamy, 2009) or time–series analyses of copula parameters (Dias and Embrechts, 2004).
The basic problem underlying diversification analyses in bear markets relates to the natural rareness of extreme events. The number of observations on which estimates are based will thus, typically, be small.

3.3 Chapter Summary

The findings on asymmetric dependence structures among different asset classes may have significant effects on portfolio decisions; this is especially the case for risk–averse investors. The usual variance–covariance approach, relying on the multivariate normal distribution, will typically overestimate diversification benefits in bear markets.

It is well–known that conditional correlations among return series vary. Although this fact, per se, does not rule out a Gaussian dependence structure, the form of such exceedance correlations conditional on returns being above or below a threshold do not correspond to the multivariate normal distribution. This may be explained by assuming that the underlying (Gaussian) dependence structure changes itself over time and is, therefore, regime–dependent. We have taken a different view and fitted several copulas allowing for dependence structures other than the Gaussian. We find that an asymmetric dependence structure of returns which can be reproduced by Archimedean copulas; furthermore, we find evidence for tail dependence in the lower tail of the return distributions.

An approach to risk management which exclusively focuses on linear correlation is not able to detect such nonlinear, asymmetric structures and may therefore overestimate diversification benefits in times of crises substantially.

Exceedance correlations are a useful tool in order to determine the potential underestimation of downside risk. Comparing empirically observed exceedance correlations to those implied by the theoretical model should always be part of an assessment of dependencies. However, in contrast to copulas, exceedance correlations are affected by the marginal distributions and are not able to tell us anything about dependencies among exceedances of unobserved extremes. Therefore, judging the appropriateness of a dependency model by exclusively trying to reproduce exceedance correlation patterns might be misleading.
In order to take investment decisions as prudent as possible, these should thus also be based on copulas allowing for lower–tail dependence.

We find that dependence structures vary across the asset class combinations considered. In general, one should not rely on only dependency concept. The Gaussian copula, for example, may be useful to model return movements in certain areas of the distribution; (asymmetric) copulas may then support analyses in the tails. But even such mixtures of copulas do not represent a one–size–fits–all approach. For example, Gao and Zhou (2010)—using daily prices of internationally traded securities—recently find evidence that some market may exhibit tail dependence, while others don't.

Chapter 4

Quantifying Credit Risk

4.1 The Credit Risk Challenge

Credit Risk is defined as the risk of losses resulting from the failure of a counterparty to fulfill its contractual obligations. Unlike market risk, which results from movements in market prices and related quantities, credit risk can result from an overdue repayment on a loan or bond, due to a counterparty having liquidity problems or even going bankrupt as happened, for example, to Enron or WorldCom.

Credit risk is typically decomposed into three components, each of which needs to be modeled. Firstly, the probability of default needs to be quantified. Secondly, the loss occurring in the case of a default (Loss Given Default, LGD) is an important ingredient to the loss distribution. Finally, the credit exposure reflects the market value of the claim and, thus, involves both market and credit risk.

However, this classical partitioning of credit risk neglects the issue of dependence modeling. In fact, default probabilities, LGD and recovery rates can be used on a granular level to monitor and assess credit risks exposure by exposure (that is, counterparty by counterparty). However, in cases of credit portfolios, such an approach may ignore relationships among defaults within the portfolio which are able to drive the tail of an aggregate loss distribution to a large extent. Just as in the market risk framework, the question of diversification effects has to be answered. Credit counterparties are typically prone to common influences, which should be explicitly modeled in order to obtain a picture as realistic as possible.

A special challenge within this setup consists of so–called low default portfolios. Our aim is therefore to assess effects of different dependency models on aggregate risk within the framework of highly–rated counterparties. Here, defaults are rare events,

which makes risk–capital estimation a difficult task. Firstly, without observations of defaults, it is not clear how to obtain an estimate of the probability of default. We do not treat this problem explicitly, and implicitly assume that such default rates are given, for example, from rating agencies. Secondly, simulation of rare events poses a computational challenge in itself.

After introducing the basic concepts of portfolio credit risk, we introduce the latent–variable model based on the multivariate normal distribution, which constitutes the most popular approach to inducing dependence among defaults. We then go on to consider alternative setups to dependency modeling. In a simulation study, we assess the effects of dependencies—and the differences among models—on defaults, and subsequently, on risk–capital estimates. The issue of reducing the computational burden in rare–event simulation is tackled by proposing an alternative method based on quasi–random sequences.

Portfolio Credit Risk

We define the credit loss of a portfolio of counterparties $i = 1, \dots n$ as

$$L = \sum_{i=1}^{n} Y_i E_i (1 - f_i) \tag{4.1}$$

where $Y_i \sim \mathrm{Ber}(\pi_i)$ is the default indicator for counterparty i: $Y_i = 1$ ($Y_i = 0$) implies default (no default). Its moments are given by $\mathrm{E}[Y_i] = \pi_i$ and $\mathrm{Var}[Y_i] = \pi_i(1 - \pi_i)$. Independently of any assumption of dependencies among obligors, the expected credit loss is given by

$$\mathrm{E}[L] = \sum_{i=1}^{n} \pi_i E_i (1 - f_i) \ .$$

If we consider two counterparties, $i = 1, 2$, we can write down the correlation between default indicators as

$$\mathrm{Corr}[Y_1, Y_2] = \rho_{1,2} = \frac{\mathrm{Cov}[Y_1, Y_2]}{\sqrt{\mathrm{Var}[Y_1]\,\mathrm{Var}[Y_2]}} = \frac{\Pr[Y_1 = 1, Y_2 = 1] - \pi_1 \pi_2}{\sqrt{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)}} \ . \tag{4.2}$$

This implies that the joint default probability can be written as

$$\Pr[Y_1 = 1, Y_2 = 1] = \pi_1 \pi_2 + \rho_{1,2} \sqrt{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)} \ , \tag{4.3}$$

and conditional probabilities are given by

$$\Pr[Y_i = 1 | Y_j = 1] = \pi_i + \frac{\rho_{1,2}}{\pi_j} \sqrt{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)} \ , \quad i, j = 1, 2 \ , \quad i \neq j \ . \tag{4.4}$$

		Counterparty 2		
		Default	No Default	
Counterparty 1	Default	0.012	0.038	0.05
	No Default	0.0380	0.9120	0.95
		0.05	0.95	1

Table 4.1: Contingency Table for $\rho_{1,2} = 0.2$

Example 4.1. *If we assume that* $\pi_1 = \pi_2 = 0.05$*,* $E_1 = E_2 = 100$ *and* $\rho_{1,2} = 0.2$*, we obtain the following contingency table of default probabilities:*
Changing to a higher correlation of $\rho_{1,2} = 0.7$*, we obtain Table 4.2. Using these tables and (squared) exposures, we an calculate the moments of the aggregate loss distribution.*

		Counterparty 2		
		Default	No Default	
Counterparty 1	Default	0.0358	0.0143	0.05
	No Default	0.0143	0.9358	0.95
		0.05	0.95	1

Table 4.2: Contingency Table for $\rho_{1,2} = 0.7$

While the expected loss equals $E[L] = 10$ *in both cases, the variance of the portfolio loss* $Var[L]$ *is influenced by correlation. In the case of* $\rho_{1,2} = 0.2$*, we obtain* $E[L^2] = 1{,}240$ *and thus* $Var[L] = 1{,}240 - 100 = 1{,}140$*, and for* $\rho_{1,2} = 0.7$ *we have* $E[L^2] = 1{,}715$ *and thus* $Var[L] = 1{,}718 - 100 = 1{,}618$*.*

This example illustrates the fact that in general a higher correlation among loss indicators leads to a higher dispersion of losses. This is why we expect risk–capital estimates to increase with correlation, a perception which will be analyzed in a simulation study. It remains to specify how correlation among the binary default indicators can be induced in an elegant way.

Latent–Variable Models for Credit Portfolios

The idea common to all latent–variable specifications is that there exists a second layer of—possibly observable—latent variables which drive the discrete counting process for the observed loss occurrences. Formally, a latent–variable model (LVM) can be defined as follows, cf. McNeil et al. (2005).

Definition 2. (Latent–variable Model) *Let $X = (X_1, \ldots, X_n)'$ be a random vector and $D \in \mathbb{R}^{n \times m}$ a deterministic matrix with elements d_{ij}. Suppose that*

$$S_i = j \quad \Leftrightarrow \quad d_{ij} < X_i < d_{i,j+1} \ , \qquad i \in \{1, \ldots, n\}, j \in \{0, \ldots, m\},$$

where $d_{i0} = -\infty$, $d_{i,m+1} = \infty$. Then, (X, D) is a latent–variable model *for the state vector $S = (S_1, \ldots, S_m)'$, where X_i are the latent variables and d_{ij} the thresholds of the LVM.*

We connect this setup to the binary default indicators by assuming that

$$Y_i = 1 \Leftrightarrow S_i = 0 \qquad \text{and} \qquad Y_i = 0 \Leftrightarrow S_i > 0 \ , \tag{4.5}$$

to indicate the "occurrence" and "non–occurrence" of a default event, i.e., we only distinguish between these two states. The probability of occurrence for process i, π_i, is now defined by

$$\pi_i = \Pr[Y_i = 1] = \Pr[X_i \leq d_{i1}] \ ;$$

a default thus occurs if the continuous, latent variable X_i falls below threshold d_{i1}.

This binary choice model is known as a structural model of default and basically compatible with common industry credit risk models based on the Merton (1974) approach, such as the KMV model. In these models, the latent variable is interpreted as the value of the obligor's assets. If their value falls below the so–called default boundary, the obligor defaults. This threshold can be expressed as π–quantile of the distribution of X_i: For normally distributed latent variables, the unconditional default probability is given by $\pi = \Phi(d_{i1})$.

In this framework, dependencies can be elegantly introduced by assuming that the obligors' asset values depend on a common factor Ψ, representing the state of the economy. Adding an idiosyncratic component of asset values, ε_i, one can write

$$X_i = \sqrt{\rho_X}\Psi + \sqrt{1 - \rho_X}\varepsilon_i \ . \tag{4.6}$$

The conditional default probability

$$p(\psi) = \Pr[X_i \leq d_{i1} | \Psi = \psi]$$

depends on asset value given a realization of the common factor. From Equation (4.6), it follows that

$$X_i | \Psi = \psi \sim \mathrm{N}(\sqrt{\rho_X}\Psi, (1 - \rho_X)) \ . \tag{4.7}$$

The realization of the common factor thus shifts the asset value distribution. Alternatively, effects of the common factor can be illustrated in terms of the distribution of standardized asset values

$$\varepsilon_i = \frac{X_i - \sqrt{\rho_X}\Psi}{\sqrt{1-\rho_X}} \sim \mathrm{N}(0,1) \ . \tag{4.8}$$

Here, different values of the common factor lead to different default thresholds, which can be written in terms of the conditional default probability from Equation (4.8) as

$$\Phi^{-1}(p(\psi)) = \frac{d_i - \sqrt{\rho_X}\Psi}{\sqrt{1-\rho_X}} \tag{4.9}$$

in the case of a normally distributed factor Ψ. The effect of changes in the common factor can, thus, be seen either as shifts in the (conditional) distribution of asset values—with fixed default probability π—or as changes in the default threshold, $p(\psi)$ for the standardized, fixed asset value distribution. This is illustrated in Figure 4.1, which shows the effect of a change in Ψ. On the left, conditional distributions of the asset value X_i are shown for different values of Ψ. As the default threshold—assuming an unconditional default probability of π=0.05—remains fixed, the shift to the right implied by an increase in Ψ leads to a lower conditional default probability, given by the area to the left of $\Phi^{-1}(0.05)$. In terms of the standardized asset value distribution, which is shown on the right, an increase in Ψ shifts the default thresholds to the left. Obviously, the results in terms of conditional default probabilities are identical.

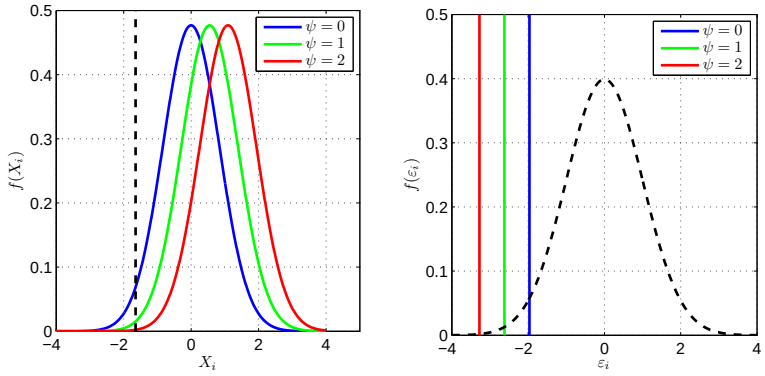


Figure 4.1: Effect of the Common Factor Ψ

The effect of asset correlation, ρ_X, on the conditional default probability depends on the value of the common factor as well. Figure 4.2 illustrates this. As can be seen from Equation 4.7, the sign of ψ determines the direction of the shift in the asset value distribution: An increase in correlation shifts the conditional asset value distribution to the right (left) for positive (negative) ψ. At the same time, correlation changes affect the variance of the distribution: The higher latent correlation, the narrower the distribution.

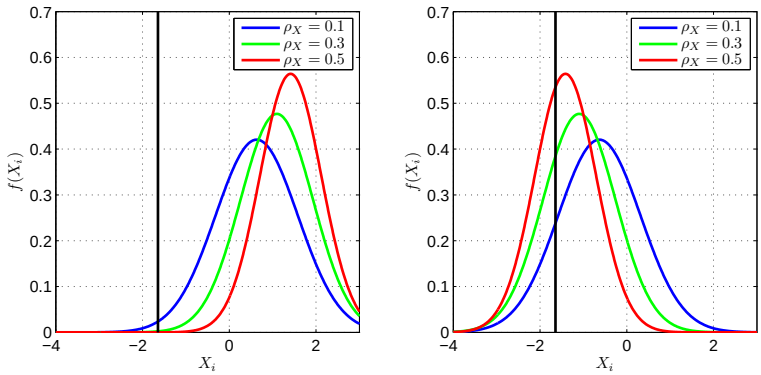


Figure 4.2: Effect of Correlation on the Asset Value Distribution

This is also illustrated in Figure 4.3 which shows conditional default probabilities for different values of ρ, depending on the common factor. Obviously, without knowledge of ψ it is not clear if an increase in correlation will lead to an increase or a decrease in the conditional default probability.

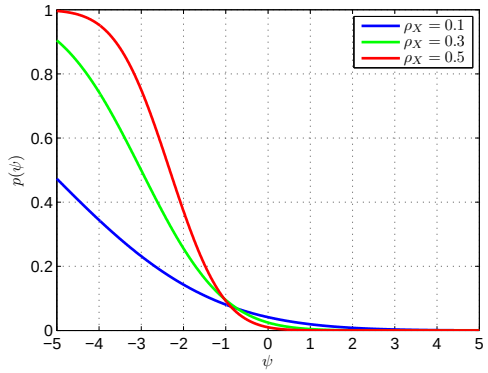


Figure 4.3: Effect of Correlation on the Conditional Default Probability

In contrast to the unobserved asset correlation ρ_X, one can observe the correlation among default indicators Y_i. This correlation is given by (see, e.g. McNeil et al., 2005, p. 344)

$$\rho_Y = \frac{\mathrm{Cov}[Y_i, Y_j]}{\sqrt{\mathrm{Var}[Y_i] \cdot \mathrm{Var}[Y_j]}} = \frac{\mathrm{E}[Y_i Y_j] - \pi_i \pi_j}{\sqrt{\pi_i(1-\pi_i)\pi_j(1-\pi_j)}}\,, \tag{4.10}$$

From (4.10) it follows that observed correlations depend on marginal occurrence probabilities, π_i and π_j, and on latent correlation, ρ_X, the latter entering via $\mathrm{E}[Y_i Y_j]$. Relationship (4.10) has been studied in detail by Gersbach and Lipponer (2003) and Foulcher et al. (2005) and is illustrated in Figure 4.4 for different values of $\pi_i = \pi_j = \pi$, assuming a normal distribution of asset values as before.

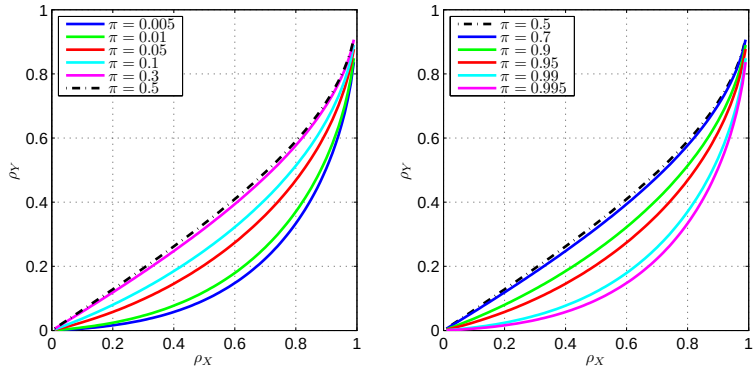


Figure 4.4: Latent versus Observed Correlation

Figure 4.4 illustrates that observed correlation, ρ_Y, is always lower than latent correlation, ρ_X, reaching its maximum value

$$\rho_Y^{\max} = \frac{2}{\pi} \arcsin \rho_X \tag{4.11}$$

at $\pi = 0.5$ (see Gersbach and Lipponer, 2003). In fact, for low default probabilities such as $\pi = 0.05\%$, a strong latent correlation may lead to a low observed correlation. In other words, latent correlation may need to be very strong in order to obtain substantial correlation among default indicators.

4.2 Estimating Risk Capital for Correlated, Rare Events

We want to assess the effects of correlated, rare credit defaults on portfolio risk as measured by VaR and ES. Additionally, we want to leave the setup of multivariate normality and assess the behavior of risk capital estimates in cases where the distribution of latent variables is non–normal. In order to elegantly solve this task, the latent variable model should be restated as Bernoulli mixture model.

Mixture models can arise when distributional parameters do not remain constant. A formal definition of a special mixture model in the spirit of McNeil et al. (2005) is as follows.

Definition 3. (Bernoulli Mixture Model) *Let* $Y = (Y_1, \ldots, Y_n)'$ *be a random vector in* $\{0,1\}^n$ *and* $\Psi = (\Psi_1, \ldots, \Psi_p)'$, $p < n$, *be a factor vector. Then,* Y *follows a* Bernoulli mixture model *with factor vector* Ψ *if there exist functions* $p_i : \mathbb{R}^p \to [0,1]$ *such that conditional on* Ψ *the elements of* Y *are independent Bernoulli random variables with* $Pr[Y_i = 1|\Psi = \psi] = p_i(\psi)$.

To keep the setup simple, we examine only exchangeable mixture models, where conditional probabilities of event occurrence are identical, i.e., $p_i(\psi) = p(\psi)$. Defining the new random variable $Q = p(\Psi)$, the observed correlation between indicator variables can then be obtained from

$$\rho_Y = \frac{\pi_2 - \pi^2}{\pi - \pi^2} \ ,$$

where $\pi = \mathrm{E}[Q]$ and $\pi_k = \mathrm{E}\left[Q^k\right]$, the latter resulting from conditional independence.[1] In fact, LVMs and Bernoulli mixture models can be viewed as two different representations of the same underlying mechanism. The following lemma from Frey and McNeil (2003) states the condition for such an equivalence.

Lemma 1. *Let (X, D) be an LVM with n-dimensional random vector X. If X has a p-dimensional conditional independence structure with conditioning variable Ψ, the default indicators $Y_i = \mathcal{I}_{X_i \leq d_{i1}}$ follow a Bernoulli mixture model with conditional event probabilities $p_i(\psi) = Pr[X_i \leq d_{i1} | \Psi = \psi]$.*

In order to generalize the factor model .. and be able to modify distributional assumptions, we restate it as

$$X_i = \sqrt{W} Z_i \ , \tag{4.12a}$$

$$Z_i = \sqrt{\rho_X}\Psi + \sqrt{1 - \rho_X}\varepsilon_i \ , \tag{4.12b}$$

where W is a scalar random variable independent of Z, so that X_i is a normal–variance mixture. Via the model (4.12), we obtain a conditional independence structure for X, which allows us to proceed using the equivalent mixture model representation.

Below, we do not only want to analyze the effect of correlation on risk–capital estimates for rare events, but also to assess consequences in terms of model risk, i.e., the use of a model which does not accurately describe the processes in place. Therefore, we compare the impact of different distributional assumptions that are frequently adopted in the credit–risk framework. To do this systematically, we calibrate the mixture models with the help of latent and observed correlations. The simplest and most common LVM, serving here as benchmark model, assumes multivariate normal latent variables. In this case, the distribution of X_i in (4.12a) corresponds to that of Z_i, implying that $W = 1$.

[1] Due to conditional independence, one obtains

$$\pi_k = \Pr[Y_1 = 1, \ldots, Y_k = 1] = \mathrm{E}[\mathrm{E}[Y_1 \cdots Y_k | Q]] = \mathrm{E}[\mathrm{E}[Y_1|Q] \cdots \mathrm{E}[Y_k|Q]] = \mathrm{E}\left[Q^k\right] \ .$$

Standardizing X_i and inserting the threshold expression $d_i = \Phi^{-1}(\pi)$, one obtains the conditional default probability

$$p(\psi) = \Pr[X_i \leq d_i | \Psi = \psi] = \Phi\left(\frac{\Phi^{-1}(\pi) - \sqrt{\rho_X}\psi}{\sqrt{1-\rho_X}}\right) . \tag{4.13}$$

In order to assess model risk, we adapt modifications suggested and analyzed in the context of credit portfolios (cf. Frey and McNeil, 2002, 2003). The first generalization of our benchmark model allows for tail dependence and a fat–tailed multivariate distribution for the latent variables by assuming a multivariate Student-t distribution. In this case, $\rho_X = 0$ means that the latent variables are uncorrelated, but they are no longer independent. The degree–of–freedom parameter of the t distribution, ν, adds flexibility and allows to control the degree of fat–tailedness of the latent variables. The setup given in (4.12) remains valid, but for X_i to have a Student-t distribution with ν degrees of freedom, the normally distributed random variable Z_i has to be multiplied by an inverse–gamma distributed random variable; i.e., $W \sim \text{InvGam}(\nu/2, \nu/2)$. The resulting conditional occurrence–probability is

$$p(\psi) = \Pr[X_i \leq d_i | \Psi = \psi] = \Phi\left(\frac{t_\nu^{-1}(\pi)W^{-1/2} - \sqrt{\rho_X}\psi}{\sqrt{1-\rho_X}}\right) , \tag{4.14}$$

where $t_\nu^{-1}(\cdot)$ denotes the inverse of the t distribution with ν degrees of freedom.

Leaving the framework defined by (4.12), two additional modifications are considered. First, we assume a Beta distribution for the mixing variable. It is a natural extension, because the interval $[0, 1]$ is the domain, and therefore allows us to interpret the mixing variable as conditional probability. Moreover, it leads to an analytically tractable model. We therefore consider a mixing variable $p(\psi) = Q \sim \text{Beta}(a, b)$. As the moments of a Beta distribution can be directly calculated from the distributional parameters, a and b, we can easily derive unconditional occurrence–probabilities[2],

$$\pi_k = \frac{\beta(a+k, b)}{\beta(a, b)} = \prod_{j=0}^{k-1} \frac{a+j}{a+b+j} , \tag{4.15}$$

and the observed correlation

$$\rho_Y = \frac{1}{a+b+1} .$$

[2]Here, we are making use of the representation of the Beta function, $\beta(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$, in terms of the Gamma function, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$, which is

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} .$$

Equation (4.15) then follows from the general recursion $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

Below, these relationships are used for model calibration.

Finally, we also consider an Archimedean copula

$$C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \ldots + \phi(u_d))$$

for the latent variables, where ϕ (the "generator") refers to the the inverse of the Laplace transform of cumulative distribution function G on $\mathbb{R}$. This choice of copula, as pointed out by Frey and McNeil (2003), is mainly motivated by its simplicity for calibration and simulation. Furthermore, Archimedean copulas allow—in contrast to the Gaussian copula—for tail dependence, the exact form of which depends on the form of the generator. To simulate latent variables with an Archimedean copula, we assume a factor, Ψ, and a sequence of uniform random variables, $U_1, \ldots, U_d$, conditionally independent given Ψ, with

$$\Pr[U_i \leq u | \Psi = \psi] = \exp(-\psi\phi(u)) \;, \qquad u \in [0, 1] \;.$$

Conditional default probabilities can be calculated from

$$Q = p(\psi) = \Pr[U_i \leq \pi | \Psi = \psi] = \exp(-\psi\phi(\pi)) \;.$$

In the special case of a Clayton copula, the generator takes the form $\phi(t) = t^{-\theta} - 1$ and Ψ needs to be gamma distributed, i.e., $\Phi \sim \mathrm{Ga}(1/\theta, 1)$. The resulting copula is characterized by lower tail dependence.[3] With this, we obtain the bivariate default probability

$$\pi_2 = \phi^{-1}(\phi(\pi) + \phi(\pi)) = (2\pi^{-\theta} - 1)^{-1/\theta} \;, \qquad \theta > 0 \;, \tag{4.16}$$

which is used for calibration to the benchmark model.

4.3 Simulation

4.3.1 Simulation Setups

In order to obtain risk–capital estimates, the loss distribution, F_L, needs to be derived. As this function is not available in closed form, we construct its empirical distribution function

$$\widehat{F}_L(l) = \frac{1}{B} \sum_{b=1}^{B} \mathcal{I}_{[0,l]}(l_b) \tag{4.17}$$

[3] We thus impose tail dependence in that part of the latent variables' distribution which is relevant for loss event occurrences. The coefficient of lower tail dependence for the Clayton copula is given by $\lambda_L = 2^{-1/\theta}$, while for the upper tail, $\lambda_U = 0$ holds.

via Monte–Carlo simulation, where l_b refers to the loss from replication b and B replications are used altogether. The resulting $\text{VaR}_{1-\alpha}$ value, based on these B replications, is then given by

$$\text{VaR}_{1-\alpha} = \widehat{F}_L^{-1}(1-\alpha) \ ,$$

and that for $\text{ES}_{1-\alpha}$ by

$$\text{ES}_{1-\alpha} = \frac{1}{B}\sum_{b=1}^{B} l_b \, \mathcal{I}_{[\text{VaR},\infty]}(l_b) \ .$$

In the benchmark model, risk capital estimates are generated as follows:

1. Set values for the observed default probabilities π and latent correlations ρ_X
2. Simulate a (standard normally distributed) factor realization, ψ, and calculate the conditional occurrence–probability $p(\psi)$ from (4.13)
3. Conduct n Bernoulli trials, using $p(\psi)$ as success probability, and sum up the number of event occurrences
4. Repeat these steps B times and calculate VaR and ES for the resulting empirical distribution.

For the distributional specifications considered, this procedure has to be modified accordingly. In order to draw from a Student-t distribution for the latent variables, we fix π and ρ_X to the same values as in the benchmark model. At the same time, we are free to set a value for the degrees–of–freedom parameter ν, a smaller value implying fatter tails for the distribution of the latent variables. In addition to drawing a factor realization from $\Psi \sim \text{N}(0,1)$, we also draw from $W \sim \text{InvGam}(\nu/2, \nu/2)$ and calculate the conditional occurrence–probability $p(\psi)$ from (4.14), which is then again used for Bernoulli trials in order to simulate loss event occurrences.

In case of a beta mixing distribution, we calculate the bivariate default probability, π_2, implied by π and ρ_X in the benchmark model.[4] Parameters a and b of the beta distribution can then be derived from π and π_2 via (4.15). Drawing from the Beta distribution with parameters a and b, we directly obtain the conditional occurrence–probability, $p(\psi)$, which is then used to conduct n Bernoulli trials. Again, the procedure is repeated B times, and values for VaR and ES are derived.

In case of a Clayton copula of latent variables, we first need to determine the value of parameter θ implied by the benchmark–model values of π and ρ_X. In order to do

[4] π_2 denotes the value of the cumulative distribution function of a bivariate normal distribution with correlation ρ_X at $\big(\Phi^{-1}(\pi), \Phi^{-1}(\pi)\big)$.

so, we again calculate the corresponding π_2 and derive θ from π and π_2 using (4.16). We then draw factor realizations from $\Psi \sim \text{Gam}(1/\theta, 1)$ which are used to obtain conditional default probabilities $p(\psi)$. Given these, we again conduct n Bernoulli trials and replicate B times in order to derive values for VaR and ES.

4.3.2 Simulation of Rare Events: Quasi–Random Numbers

The general problem which we are facing in our application is the computation of the expected value

$$\theta = \text{E}[h(l)] = \int_{-\infty}^{\infty} h(l)\text{d}F_L \ . \tag{4.18}$$

For example, to estimate the exceedance probability $\Pr[L > \alpha]$, from which VaR_{alpha} can be derived, one would set $h(l) = \mathcal{I}_{(a,\infty)}(l)$; if we wish to estimate Expected Shortfall, $\theta = \text{E}[L|L \geq \text{VaR}]$, the function $h(l)$ is of the form

$$h(l) = l\ \mathcal{I}_{[\text{VaR},\infty)}(l) \ . \tag{4.19}$$

The estimation of the integral in equation (4.18) is performed via the approximation

$$\theta \approx \int_{-\infty}^{\infty} h(l)\text{d}\widehat{F}_L \ , \tag{4.20}$$

where $\widehat{F}_L$ refers to an empirically estimated distribution function of L. This quantity is the output of simulations. In the ordinary MC setup, $\widehat{F}_L$ from equation (4.20) refers to the empirical distribution function

$$\widehat{F}_L^{\text{MC}}(l) = \frac{1}{B}\sum_{b=1}^{B} \mathcal{I}_{[0,l]}(l_b^{\text{MC}}) \ , \tag{4.21}$$

where l_b^{MC} is drawn from the appropriate distribution function.
The distinguishing feature of this basic approach is the pseudo–randomness of realizations l_b^{MC}. Assuming for simplicity that we only face one (invertible) distribution function F available in closed form, l_b^{MC} is obtained from

$$l_b^{\text{MC}} = F^{-1}(u_b^{\text{MC}}) \ , \tag{4.22}$$

using the probability–integral transform, where $U_b \sim \text{Unif}(0, 1)$. However, in our application F_L is not available in closed form. therefore, simulations are rather based on individual event occurrences y_i, which are summed up in order to obtain the total loss of replication b, $l_b = \sum_{i=1}^{n} y_i$. Nevertheless, using the inverse of the Bernoulli distribution function, the probability integral transform and all arguments which follow equally apply to simulation of y_i, $i = 1, \ldots, n$. The Monte–Carlo estimate

$$\hat{\theta}_n^{\text{MC}} = \frac{1}{B}\sum_{b=1}^{B} h(l_b^{\text{MC}}) \tag{4.23}$$

is based on B independent, F_L–distributed samples $l_1, \ldots, l_B$ and amounts to an averaging over the B pseudo–random realizations. From the Glivenko–Cantelli Theorem (see, e.g., Laha and Rohatgi, 1979, Remark 2.5.1), we know that $\widehat{F}^{\mathrm{MC}}(l) \to F(l)$ uniformly in l as $B \to \infty$. Using the Extendend Helly–Bray Theorem (see, e.g., Laha and Rohatgi, 1979, Theorem 3.1.4), we conclude that

$$\hat{\theta}_n^{\mathrm{MC}} \stackrel{B\to\infty}{\longrightarrow} \theta \; ; \tag{4.24}$$

for h is a bounded, real valued function on $\mathbb{R}$.
From Theorem 1.1. of Niederreiter (1992) we know that if h is second–power integrable, then, for any $B \geq 1$,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{B} \sum_{b=1}^{B} h(l_b^{\mathrm{MC}}) - \mathrm{E}[h] \right)^2 dF_L(l_1) \cdots dF_L(l_B) = \frac{\sigma^2(h)}{B} \; , \tag{4.25}$$

where h refers to the variance of the integrand

$$\sigma^2(h) = \int_{-\infty}^{\infty} (h - \mathrm{E}[h])^2 \, dF_L \; . \tag{4.26}$$

As (4.25) gives us the mean–square error of the MC approximation, we conclude that the error in (4.23) is on average $\sigma(h)/\sqrt{B}$ and thus of order $\mathrm{O}(B^{-1/2})$.[5] The remarkable advantage of MC simulation consists of the fact that this bound does not depend on the dimension d of the integrand—which is equal to $d = 1$ in our application—as would be the case for a multidimensional integration rule for (4.18).

Nevertheless, we know from (4.24) that we obtain a *probabilistic* bound, such that the accuracy of the MC approximation is not guaranteed. Furthermore, the error bound is valid for any square–integrable h, and no additional gain in accuracy can be obtained from further regularity of the integrand. Apart from these drawbacks, for equation (4.25) to hold, true *random* independent samples are required, which cannot be generated in reality. Besides, even if $l_1, \ldots, l_B$ were based on a random sample, this would allow for clumping effects due to the independence of the B elements.

When estimating θ via Monte–Carlo, problems arise in case of rare events.[6] For example, in order to reliably simulate losses beyond VaR_α with high α, many observations are needed. One thus faces either a high mean–square error of the MC approximation

[5] In general, if f, g are two functions defined on some subset of $\mathbb{R}$, $f(x) \in \mathrm{O}(g(x))$ as $x \to \infty$ if and only if there exists a positive real number c and a real number x_0 such that

$$|f(x)| \leq c|g(x)| \text{ for } x > x_0 \; ; \tag{4.27}$$

see, e.g., Cormen et al. (2001).

[6] A detailed account on rare event simulation can be found, for example, in Rubino and Tuffin (2009).

or a substantial computational effort. Variance reduction techniques such as Importance Sampling[7] aim at achieving a lower mean–square error of the MC approximation, $\sigma^2(h)/B$, by reducing the integrand's variance $\sigma^2(h)$. Quasi–random sequences pursue a different strategy. They increase accuracy by generating points which are too evenly distributed to be random.

The simulation of losses via quasi–random sequences is also based on the approximation (4.20) and can thus be written as

$$\hat{\theta}_n^{\text{QMC}} = \frac{1}{B}\sum_{b=1}^{B} h(l_b^{\text{QMC}}) \; . \tag{4.28}$$

The difference to the MC estimate (4.23) is the way in which the realizations l_b are obtained. While they are pseudo–random in the ordinary MC setup, they are now based on quasi–random sequences; that is,

$$l_b^{\text{QMC}} = F^{-1}(u_b^{\text{QMC}}) \; , \tag{4.29}$$

where u_b is a deterministic sequence on the unit interval characterized by its low discrepancy.

Uniformity, Discrepancy and Variation

The approximation

$$\theta = \text{E}[h] = \int_{-\infty}^{\infty} h(l) dF_L \approx \frac{1}{B}\sum_{i=1}^{B} h(F^{-1}(u_b^{\text{QMC}})) \; , \tag{4.30}$$

which is the basis for the simulation procedures described above, performs best if the points $u_1, \ldots, u_B$ are uniformly distributed over the unit interval. This can easily be seen from the fact that in the case of such a uniform distribution,

$$\lim_{B\to\infty} \frac{1}{B}\sum_{b=1}^{B} h(u_b) = \int_{-\infty}^{\infty} h(u) dF_L \tag{4.31}$$

holds. In other words, the closer the empirical distribution function of $u_1, \ldots, u_B$ to the uniform, the smaller the approximation error in (4.20). This observation is the intuitive motivation for quasi–random sequences. Clearly, only finite sequences ($B < \infty$) can be used, so that the approximation error in (4.30) cannot completely be eliminated. However, the (deviation from the) desired uniformity can be used to rank sequences with respect to the approximation errors they induce.
As the dimensionality of the integrand, d, plays an important role in the discussion

[7] An overview of variance reduction techniques is given by Glasserman (2004).

of quasi–random sequences, we will consider the general case $d \geq 1$ in the following, although in our concrete case, $d = 1$ holds. Define a set $J \in [0,1)^d$ and a sequence $\boldsymbol{u_1}, \ldots, \boldsymbol{u_B} \in [0,1)^d$. The discrepancy as a measure of deviation from uniformity can be written as

$$D(\boldsymbol{u_1}, \ldots, \boldsymbol{u_B}; J) = \sup_{J \in E} \left| \frac{1}{B} \sum_{b=1}^{B} \mathcal{I}_J(\boldsymbol{u_b}) - \lambda_d(J) \right| \tag{4.32}$$

where E refers to the collection of all subsets of $(0,1]^d$ and $\lambda_d(J)$ is the d–dimensional Lebesgue measure, assigning a volume to J. Therefore, minimizing discrepancy means that for all subsets E the fraction of points inside them are as close as possible to the subsets' volumes. Depending on the definition of these subsets, different kinds of discrepancy measures are obtained. If E in equation (4.32) is restricted to subrectangles of $[0,1)^d$ which are of the form $\prod_{j=1}^{d}[v_j, w_j)$ with $0 \leq v_j < w_j \leq 1$, one obtains the ordinary (or extreme) discrepancy, D_B; if it contains only subrectangles with one corner of zero, i.e., of the form $\prod_{j=1}^{d}[0, v_j)$, the resulting measure is called the star discrepancy D_B^*.

The discrepancy measure refers to the (quasi–random) points serving as inputs to the integrand $h(l_b) = h(F^{-1}(u_b))$. To obtain an upper bound for the approximation error, we additionally need to consider the variation of the integrand itself. For a multidimensional concept of total variation of a function h on J, let $\Delta(h, J)$ denote function values of h at the vertices of J with opposite signs at adjacent vertices. The variation in the sense of Vitali is then defined as

$$V^{(d)}(h) = \sup_{\mathcal{P}} \sum_{J \in \mathcal{P}} \Delta(h, J) \ , \tag{4.33}$$

where $\mathcal{P}$ ranges over all partitions of the unit hypercube into subintervals of the form of J. If the integrand h is smooth on $[0,1)^d$, the more convenient representation

$$V^{(d)}(h) = \int_{[0,1)^d} \left| \frac{\partial^d h}{\partial u_1 \cdots \partial u_d} \right| \mathrm{d}u_1 \cdots \mathrm{d}u_d \tag{4.34}$$

is obtained. This leads to the variation in the sense of Hardy and Krause which is given by

$$V(h) = \sum_{k=1}^{d} \sum_{1 \leq i_1 < \ldots < i_k \leq d} V^{(k)}(h; i_1, \ldots, i_k) \ . \tag{4.35}$$

For more details on variation, see, e.g., Owen (2005).
If h has bounded variation $V(h)$ on $[0,1)^d$, i.e., $V(h)$ is finite, then, for any $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_B \in [0,1)^d$ we obtain the Koksma–Hlawka inequality (Koksma, 1942/43; Hlawka, 1961))

$$\left| \frac{1}{B} \sum_{b=1}^{B} h(\boldsymbol{u_b}) - \int_{[0,1)^d} h(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \right| \leq V(h) D_B^*(\boldsymbol{u_1}, \ldots, \boldsymbol{u_B}; J) \ . \tag{4.36}$$

This inequality gives us an upper bound for the approximation error and implies that sequences characterized by a small star discrepancy D_B^* guarantee a small error in the QMC integration. It is similar to (4.25) in that it relates the error magnitude to a product of one term depending on properties of the integrand h and a second one depending on properties of the sequence. However, in contrast to the ordinary Monte Carlo case, (4.36) gives us an absolute bound.
In contrast to this "worst case" error, Wozniakowski (1991) analyzes the *average* integration error in (4.30) and shows that it is equal to an alternative discrepancy measure based on the L_2 norm; see also Albrecher et al. (2003). In fact, Morokoff and Caflisch (1994) argue that the Koksma–Hlawka inequality is a vast overestimate at least for "half–differentiable" functions (whose Hölder exponent is sufficiently close to 1/2).

For the sequence we apply in our analyses, and quasi–random sequences in general, it can be shown that

$$D_B^* \in \mathrm{O}\left(\frac{(\ln(B))^d}{B}\right) \; ; \tag{4.37}$$

see Chapter 3 of Niederreiter (1992) and Meijer (1968) for the exact expression of the upper bound. This is a substantial improvement on the probabilistic ordinary Monte Carlo error. In fact, as Niederreiter (1978) shows, the star discrepancy of a pseudo–random sequence satisfies

$$D_B^* \in \mathrm{O}\left(\left(\frac{\ln(\ln(B))}{B}\right)^{1/2}\right) \tag{4.38}$$

which is in line with the probabilistic error bound of order $\mathrm{O}(B)^{-1/2}$. The quasi–random sequence covers the unit hypercube more uniformly and thus promises a faster convergence. This is illustrated in Figure 4.5, which clearly shows the clustering occurring in the case of pseudo–random numbers.

Nevertheless, quasi–random sequences outperform pseudo–random ones only beyond a certain number of elements B, this threshold increasing with dimensionality d. As Sarkar and Prasad (1987) argue, this is due to the fact that for small B, the effect of $(\ln(B))^d$ in (4.37) dominates, while for large B, it becomes insignificant and the discrepancy decreases as B^{-1}. In other words, and as Caflisch et al. (1997) put it, the powers of $\ln(B)$ do not become negligible at any computationally possible sample size.[8] Summing up, when considering the integration of a function h with bounded variation $V(h)$, the Koksma–Hlawka inequality guides us to look for sequences with the small-

[8]However, for certain high–dimensional integrands, QMC methods still outperform MC for realistic sample sizes. Caflisch et al. (1997) explain this based on the concpet of effective dimension of the integrand, which is still low in these cases.

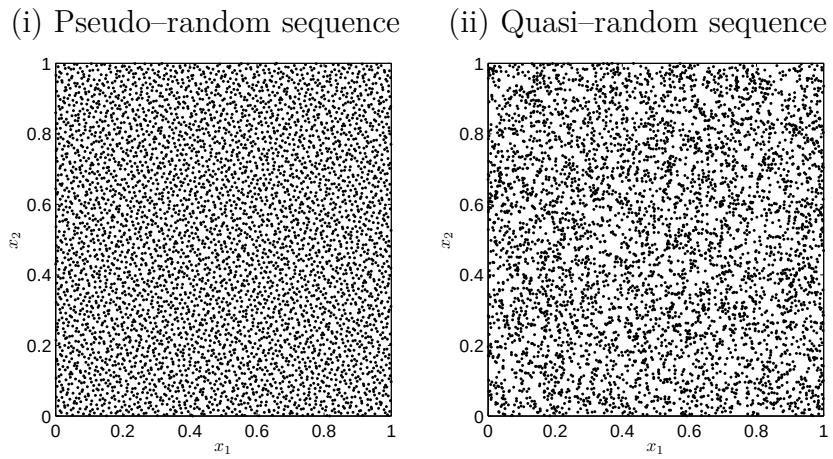


Figure 4.5: 2–dimensional Projections: Pseudo– versus Quasi–Random Sequence, $B = 5,000$

est possible star discrepancy D_B^*. This is equivalent to searching for sequences whose empirical distribution function is as close as possible to the uniform distribution. Such sequences can then outperform ordinary MC in terms of the order of magnitude of the approximation error bound.

The Van der Corput and Halton Sequences

Among the deterministic quasi–random sequences, those that have attracted most attention rely on the base-p expansion. Given a base $p \geq 2$, any integer $k > 0$ can be written as

$$k = \sum_{j=0}^{J(k)} a_j(k) p^j = (a_{J(k)} \cdots a_0)_p \tag{4.39}$$

with $a_j \in \{0, 1, \ldots, p-1\}$. The radical inverse function in base p

$$\psi_p(k) = \sum_{j=0}^{J(k)} \frac{a_j(k)}{p^{j+1}} = (.a_0 \ldots a_{J(k)})_p \tag{4.40}$$

consists of a symmetric reflection of the expansion around the decimal point and therefore leads to a mapping of k to a point in $[0, 1)$. The base-p Van der Corput sequence is then given by the sequence $\psi_p(0), \psi_p(1), \psi_p(2), \ldots$.

Example 4.2. *Taking $p = 3$ as base, we can express $k = 7$ as*

$$k = 7 = (21)_3 = 2 \cdot 3^1 + 1 \cdot 3^0 \,, \tag{4.41}$$

and therefore

$$\psi_3(7) = (.12)_3 = \frac{1}{3^1} + \frac{2}{3^2} = 0.5556 \,.$$

In the same way, we can, for example, obtain the unique base–3 expansions

$$k = 17 = (122)_3 \quad \rightarrow \quad \psi_3(17) = (.221)_3 = 0.9259 \ ,$$
$$k = 70 = (2121)_3 \quad \rightarrow \quad \psi_3(70) = (.1212)_3 = 0.6173 \ .$$

The Halton Sequence (Halton, 1960) is the first analysis and simplest construction of a low–discrepancy sequence for an arbitrary dimension d. Its coordinates follow Van der Corput sequences in distinct bases. That is, if $p_1, \ldots, p_d$ are relatively prime[9],

$$s_k = (\psi_{p_1}(k), \ldots, \psi_{p_d}(k)), \qquad k = 0, 1, 2, \ldots \tag{4.42}$$

is a Halton sequence in dimension d. The point $k = 0$ is theoretically contained in the sequence but will be omitted in the following due to the fact that for a distribution function F, $F^{-1}(\psi_p(0)) = F^{-1}(0) = -\infty$, which would cause the integrand h to be unbounded and thus lead to an inapplicability of the Koksma–Hlawka inequality.[10] Therefore, in our application, k coincides with the index referring to the number of replications, $b = 1, \ldots, B$, when using the Halton sequence for QMC simulation.
The fact that this construction leads to a higher uniformity than, for example, a grid construction, is illustrated in Figure 4.6, which shows the first ten elements of the one–dimensional Halton sequence with base $p = 3$, highlighting by arrows the elements newly added in each step $b = 1, 2, \ldots, 10$. In general, the numbers lie in circles of p increasing terms and are separated by $1/p$ within each of these circles.[11] Once a grid of $1/p^m$ has been completed, the next circle starts filling a grid at the level $1/p^{m+1}$. In this way, numbers are inserted in a maximally balanced way.

On the search for a uniformly distributed sequence one would, at a first glance, suggest a simple grid over the unit interval. Figure 4.6 clarifies why such a grid is nevertheless inferior to low–discrepancy sequences: If the function h is nearly separable in its arguments, the values at the B^d grid points contain nearly the same information as

[9] Two integers are relatively prime if they share no common positive factors (divisors) except 1. Using the notation (m, n) to denote the greatest common divisor, two integers m and n are relatively prime if $(m, n) = 1$. Relatively prime integers are sometimes also called strangers or coprime and are denoted $m \perp n$. Two distinct primes p and q are always relatively prime, $(p, q) = 1$, as are any positive integer powers of distinct primes p and q, $(p^m, q^n) = 1$.

[10] Owen (2006) argues that Halton sequences avoid regions around corners of the unit cube, so that the error bound may even hold without this restriction.
Another potential violation of the Koksma–Hlawka inequality in our application might be caused by the presence of the indicator function within the integrand h. For this specific problem, Jin and Zhang (2006) suggest a smoothing via Fourier Transforms, which can theoretically lead to a faster convergence of quasi–Monte Carlo methods. As our simulation results clearly favor the use of quasi–random sequences, we have not pursued this topic further yet.

[11] As we leave out $\psi_3(0)$, the first circle in Figure 4.6 consists only of two elements.

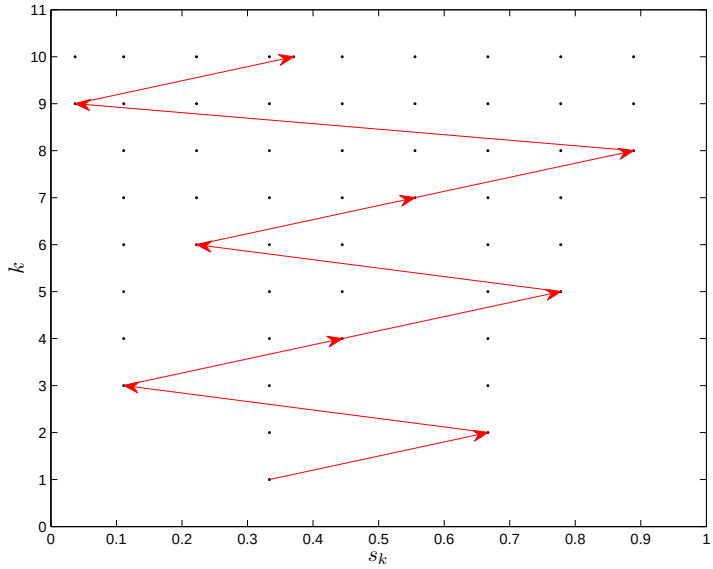


Figure 4.6: 1–dimensional Halton Sequence $s_k = \psi_3(k)$, $k = 1, \ldots, 10$

only Bd points would. Furthermore, large intervals within $[0, 1]$ are left empty, and the total number of points B has to be determined in advance. If the need for a refinement should emerge later, the grid can only be changed by increasing the number of points by a factor 2^d, which leads to a discrepancy of size $O(1)$.
Therefore, at least for small and moderate dimensions, quasi–random sequences combine the advantages of random sequences—points can be added incrementally—and those of a grid, avoiding a clumping of points.[12]

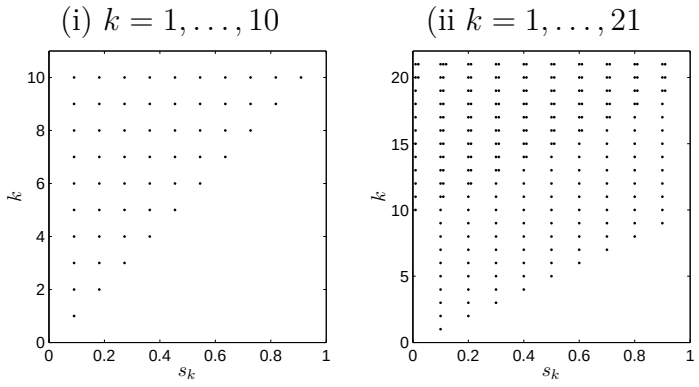


Figure 4.7: 1–dimensional Halton Sequence $s_k = \psi_{11}(k)$

The difficulties which may arise when using Halton sequences for integrands of high

[12]In applications which require an estimation of the approximation error, these deterministic sequences can additionally be randomized; see Owen (1995) for a survey on this topic.

dimensionality stem from the fact that a Halton sequence in dimension d requires d relatively prime numbers. A high dimensionality thus implies that high prime numbers have to be used. The effect of this is illustrated in Figure 4.7. It shows the one–dimensional Halton sequence with base $p = 11$ for $B = 10$ and $B = 21$. Here, the successive "filling" of the unit interval described above obviously leads to a clustering of points if the length of the sequence is not either chosen to contain exactly $p - 1$ or a very large number of points. For example, using only $b = 5$ elements, all generated points lie within the lower half of the unit interval.

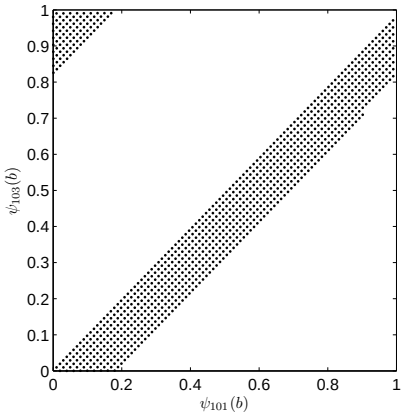


Figure 4.8: 2–dimensional Projection: Halton Sequence $(\psi_{101}(k),\psi_{103}(k))$

As an extreme, but not yet unrealistic example of the effect this non–uniformity for higher dimensions can have, Figure 4.8 shows the bivariate projection of the two–dimensional Halton sequence with bases $p_1 = 101$, $p_2 = 103$ and $B = 10,000$ points. Here, the points only cover a subspace of $[0, 1]^2$ and therefore lead to a high discrepancy. This property of Halton sequences suggests that their usefulness decreases with increasing dimension of the integration problem. For high dimension, modifications such as the leaped Halton sequence or the Reverse–Radix algorithm of Kocis and Whiten (1997) may mitigate these effects.
Our application requires a maximum of three relatively prime numbers. This is far below the dimensionalities deemed to be problematic; for example, Morokoff and Caflisch (1995) find that among the commonly used quasi–random sequences the Halton gives best results as long as $d \leq 6$. Therefore we stick to this sequence as the simplest, most intuitive one, although others, such as the sequences of Faure (1982), Sobol (1967) or Niederreiter (1992) could equally be used in our framework.

The use of quasi–random sequences reduces the computational burden immensely. The results are qualitatively identical to those obtained from a standard Monte–Carlo approach. The presentation of results in the following section is based on ordinary Monte–Carlo procedures, in order to demonstrate the effect of the number of replications on

risk–capital estimates.

4.3.3 Simulation Results

For each of the models discussed above we simulate defaults and estimate risk capital for different levels of latent correlation, ρ_X. In doing this, we use the multivariate normal LVM as benchmark model to which we calibrate the other models. Risk will be measured by the VaR and ES of the distribution of the number of losses within the portfolio.[13]

In a first simulation study, we assume $n = 1,000$ loss processes, to match the setup of the study in Frey and McNeil (2001). Specifying $B = 100,000$ replications, we observe for all models a counterintuitive behavior of the VaR values when default probabilities are low ($\pi \leq 0.01$): Starting from perfect correlation, VaR values increase as correlations decrease. We consider confidence levels between 95% and 99.9% and find that this effect is more pronounced for lower confidence levels, i.e., when decreasing ρ_X, we observe it first for $\text{VaR}_{0.95}$, then for $\text{VaR}_{0.99}$, and only for very low default probabilities ($\pi \leq 0.0001$), $\text{VaR}_{0.999}$ is also affected. An illustration of this phenomenon is given in Figure 4.9, which plots the 99% VaR (on a logarithmic scale) as a function of the level of latent correlation and default probability, π.

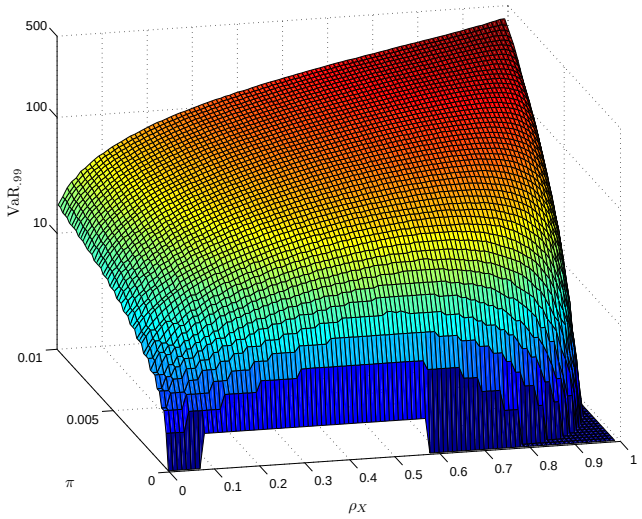


Figure 4.9: Simulated $\text{VaR}_{0.99}$ figures from a Bernoulli mixture model with multivariate normal latent variables, $\pi \in [1.0\text{e-}003, 0.01]$

For $\pi = 0.01$, VaR behaves as expected: It increases in ρ_X over the entire range of latent correlations. However, for lower levels of π, such as $\pi = 0.005$, VaR decreases

[13] This amounts to assuming a fixed exposure and thus a concentration on default risk.

with increasing correlation above a certain threshold of ρ_X; the lower π, the lower this threshold value. This effect is the more pronounced, the fatter the tails of the distribution of latent variables. This is shown in Figure 4.10, where an intermediate occurrence–probability of $\pi = 0.001$ is held fixed. For $\nu = 100$, VaR grows as the

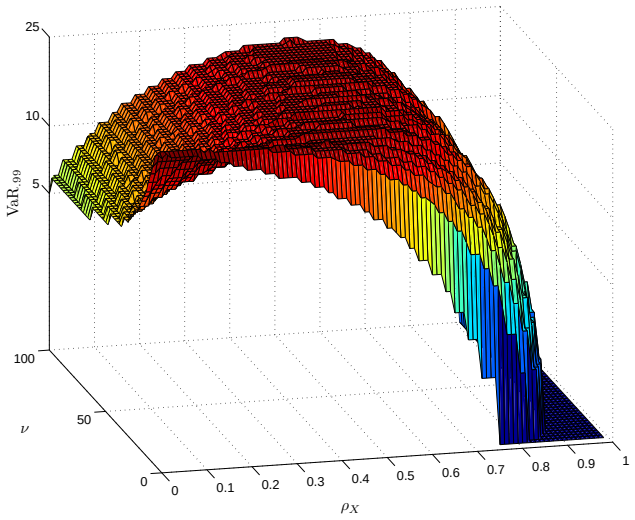


Figure 4.10: Simulated $\text{VaR}_{0.99}$: Bernoulli Mixture Model, Multivariate t–distributed Latent Variables, $\pi = 0.001$, $\nu \in [4, 100]$

latent correlation increases to $\rho_X \approx 0.5$ and decreases for higher levels. The lower ν, the broader the range of ρ_X for which this peculiar behavior occurs. For $\nu = 4$, VaR values decrease over the entire range of latent correlations, ρ_X.
For ES, using 100,000 replications, we obtain ambiguous results. Figure 4.11 shows a behavior which corresponds to intuition, i.e., ES increases with correlation. However, setting a different seed in the simulation can result in ES–increases as correlation decreases when the default probability is very low ($\pi \leq$ 1.0e-005).

The results presented so far are based on 100,000 replications. In order to assess the uncertainty in risk capital estimates, we modify the simulation setup and calculate several risk capital estimates for each combination of ρ_X and π, increasing the number of replications up to $B = 10,000,000$. In order to avoid predominance of risk capital estimates of zero, we now set the number of loss processes to $n = 100,000$ and concentrate on a confidence level of 99.9%.[14] Due to the substantial computational burden of this task, we hold the default probability fixed at $\pi =$ 1.0e-005. To assess convergence and

[14]The results from our simulation study can be reproduced for other combinations of confidence levels and default probabilities; as described, the counterintuitive behavior of VaR also occurs for lower confidence levels.

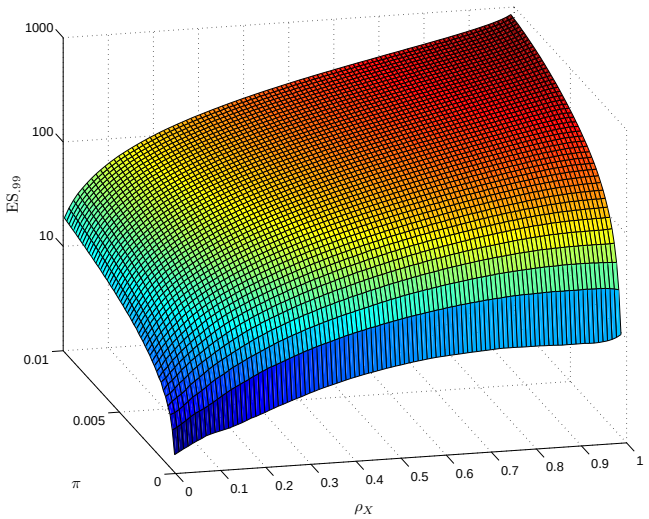


Figure 4.11: Simulated $ES_{0.99}$: Bernoulli Mixture Model, Multivariate Normal Latent Variables, $\pi \in [1.0\text{e-}003, 0.01]$

reliability of risk capital estimates, we conduct 250 simulations for each ρ_X. Figures 4.12 to 4.16 compare boxplots of these 250 risk capital estimates, where we replace the median by the mean over the 250 estimates for assessing convergence.

Figure 4.12 confirms the findings of our first simulations: Above a certain level of latent correlation ($\rho_X \approx 0.5$), VaR estimates may increase with decreasing latent correlation. The dispersion of the 250 risk capital figures is the highest for medium levels of correlation. As was to be expected, it decreases with an increasing number of replications; however, VaR still increases with decreasing correlations in the upper region of ρ_X. In view of these observations, we conclude that this counterintuitive behavior is not due to convergence issues and cannot be eliminated by increasing the number of replications. It appears that $B = 1,000,000$ replications are sufficient to obtain reliable VaR figures. Therefore, we restrict our attention to this value of B when analyzing VaR behavior in the following.

Figure 4.13 shows the ES figures resulting from the very same simulations. It becomes evident that the risk–capital estimates may, just like in the VaR case, rise with decreasing correlations; but in the case of ES figures, this is not true for their mean. Therefore, increasing the number of replications to $B^{\max} = 10,000,000$ substantially reduces the occurrence frequency of ES–increases caused by correlation declines. The restriction to $B^{\max}$ is due to computational limitations.

The results from the previous study regarding the fat–tailedness of the latent variables' distribution are confirmed by Figure 4.14. Just as before, VaR figures decrease in ρ_X

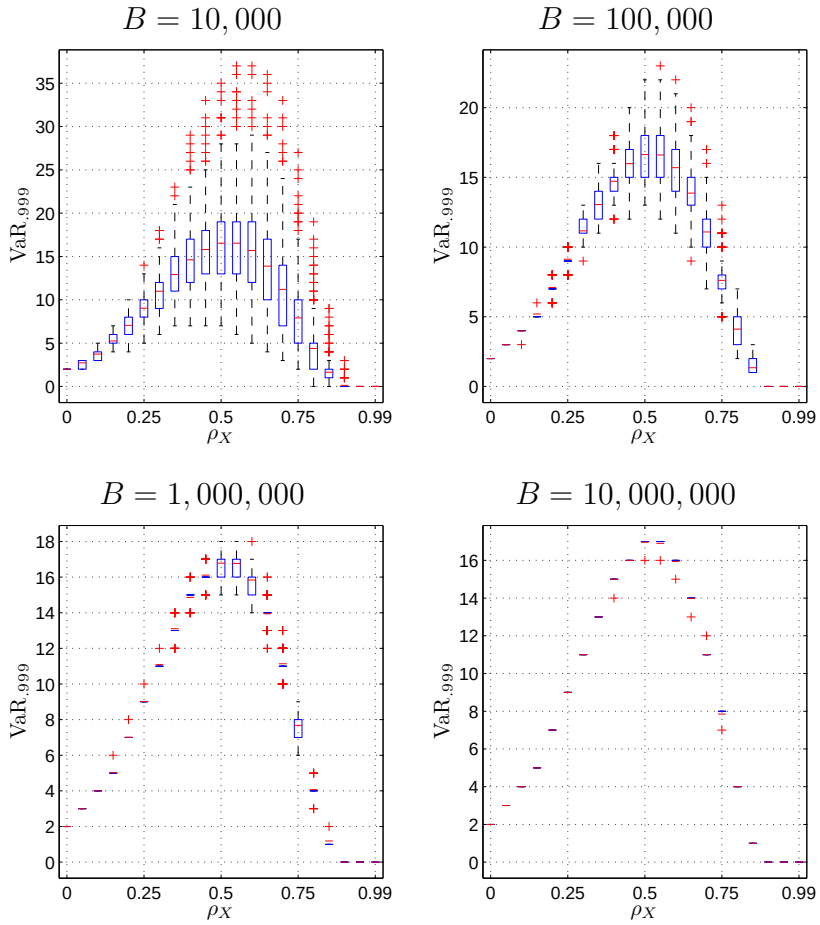


Figure 4.12: Mean–boxplots of $\mathrm{VaR}_{0.999}$: Bernoulli Mixture Model, Multivariate Normal Latent Variables, $\pi = 1.0\text{e-}005$

over the entire range of latent correlations. In addition, we observe that the dispersion of the ES–values is much higher than in the case of normally distributed latent variables, implying that ES–increases caused by correlation declines can be observed more often. Figures 4.15 and 4.16 confirm the previous findings for the Beta mixing distribution and the Clayton copula of latent variables.

4.4 Interpretation of Results

The simulation study described in the previous section is conducted in order to find out whether a general rule for the effects of (less than perfect) correlation between operational loss processes can be found. We introduce dependence through latent variables and assume that the losses are modeled jointly before estimating risk capital, i.e., we do not assess subadditivity properties.

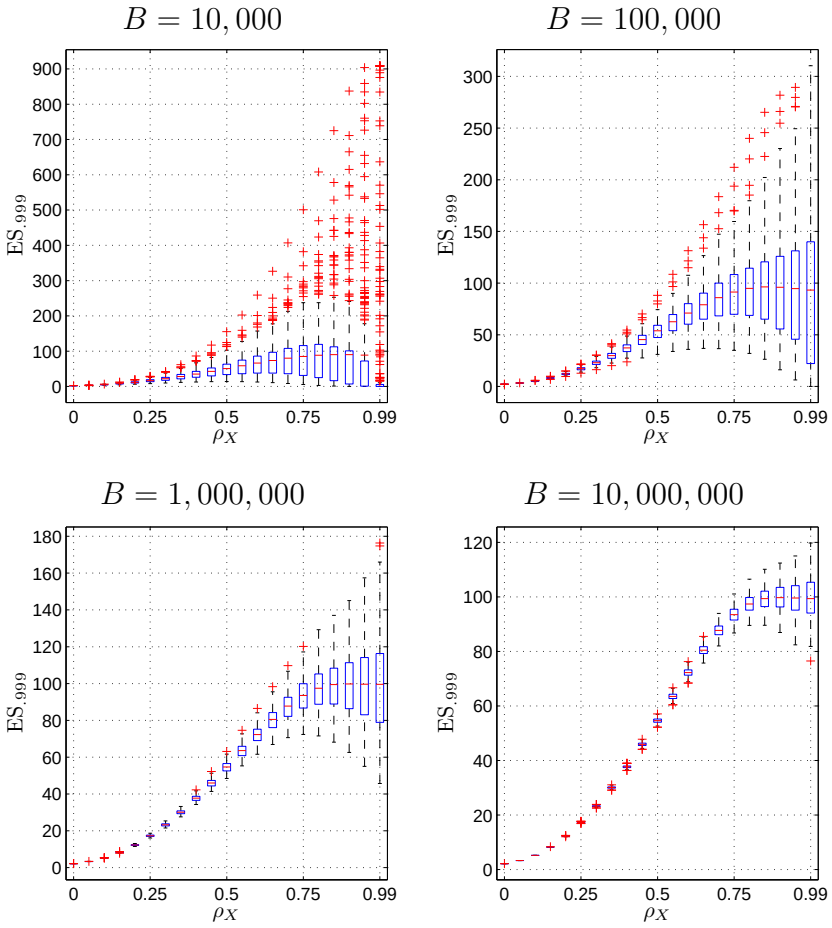


Figure 4.13: Mean–boxplots of $ES_{0.999}$: Bernoulli Mixture Model, Multivariate Normal Latent Variables, $\pi = 1.0\text{e-}005$

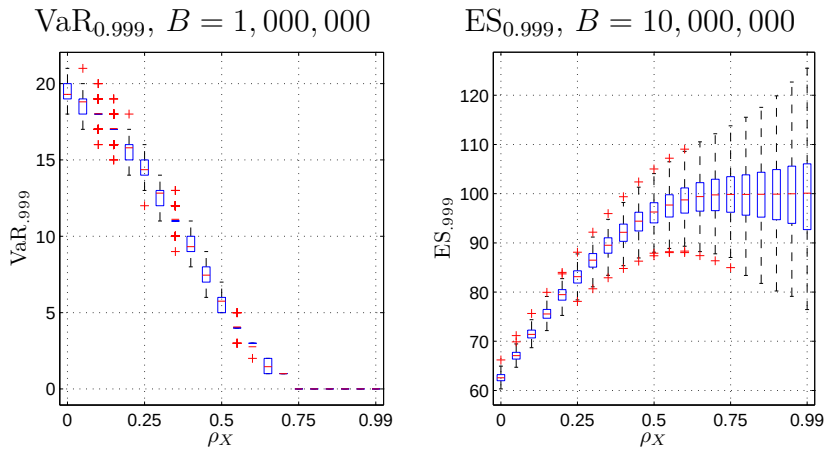


Figure 4.14: Mean–boxplots of $VaR_{0.999}$ and $ES_{0.999}$: Bernoulli Mixture Model, Multivariate t–distributed Latent Variables, $\nu = 4$, $\pi = 1.0\text{e-}005$

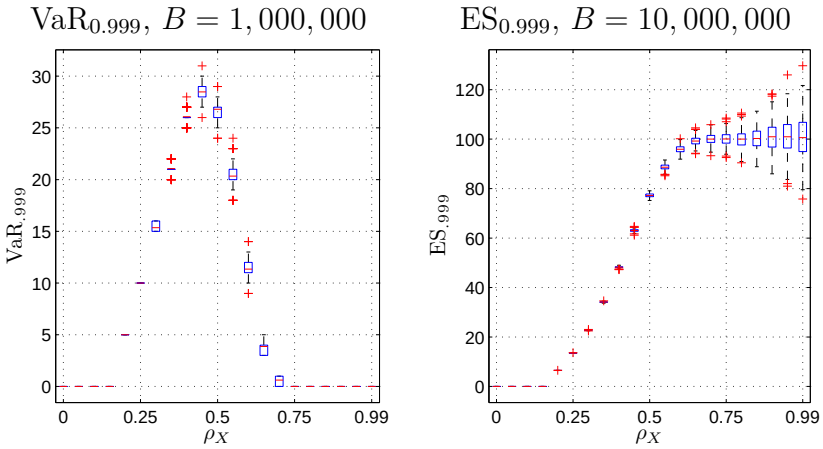


Figure 4.15: Mean–boxplots of $\text{VaR}_{0.999}$ and $\text{ES}_{0.999}$: Bernoulli Mixture Model, Beta Mixing Distribution, π = 1.0e-005

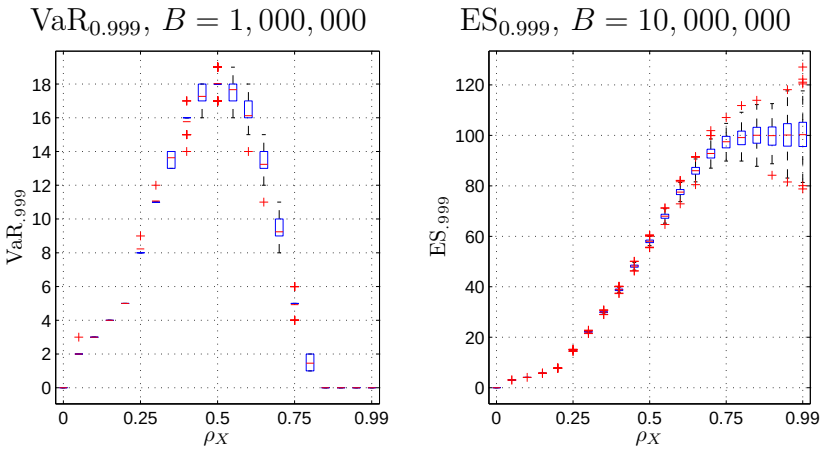


Figure 4.16: Mean–boxplots of $\text{VaR}_{0.999}$ and $\text{ES}_{0.999}$: Bernoulli Mixture Model with Clayton Copula, π = 1.0e-005

Surprisingly, we find that for all distributional assumptions considered, risk capital estimates do not necessarily decrease when departing from the assumption of perfect positive correlation. This counterintuitive effect is observed for low default probabilities, i.e., for rare events ($\pi \leq 0.01$), and only above a certain value of ρ_X, this threshold decreasing with increasing tail dependence of the latent variables' distribution. Therefore, the effect is of practical relevance only for certain "extreme" situations, which can be summarized as risk capital estimation at a very high confidence level of very rare, but (possibly tail–)dependent event occurrences.

The explanation for the unexpected behavior of risk capital estimates with respect to

different correlation assumptions is illustrated in Figure 4.17. It shows 10,000 draws from a bivariate normal distribution for two different correlation assumptions. The solid line represents the thresholds implied by the default probability of $\pi = 0.01$. In the upper plot, where the latent correlation is set to $\rho_X = 0.1$, this threshold leads to four joint "occurrences" (in the southwestern quadrant) and 9,798 joint "non–occurrences". In the lower plot with correlation $\rho_X = 0.9$, the concentration on extremes leads to 94 joint "occurrences" and 9,854 joint "non–occurrences". As it turns out, high correlation not only leads to more events, but also to more joint "non–events". It is this phenomenon which moves Value–at–Risk towards zero as correlation levels rise.

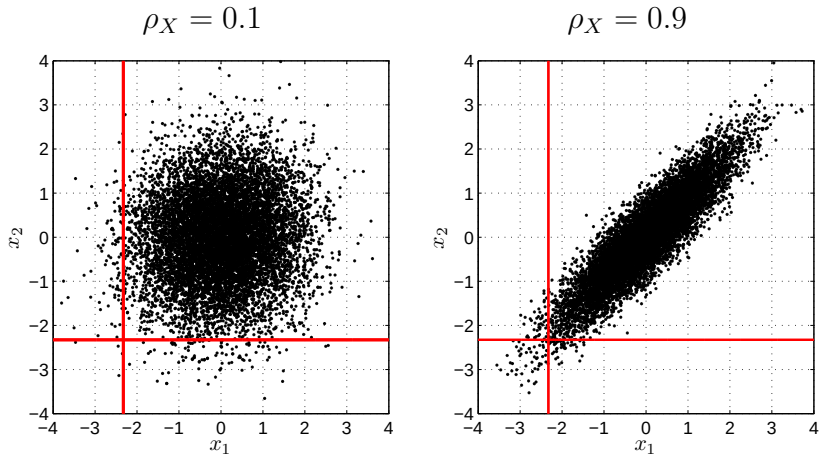


Figure 4.17: Scatterplots of Bivariate Normal Distributions

In our simulation study, we keep the setup as simple as possible. That is, we assume exchangeable latent variables and base risk capital estimates on constant latent correlations and default probabilities. In reality, these assumptions will not be met: Default probabilities change over time, e.g., due to changes in the economic environment. A dynamical consideration could therefore provide more insight on the practical implications of our findings.

4.5 Chapter Summary

In the case of low–default portfolios, as, foe example, in the case of highly–rated counterparties or very short–term exposures, there is a need to simulate rare events in order to obtain a risk–capital estimate.

Our results show that it is very important to assess the impact of correlations within the chosen modeling framework. In the case of rare events, simulated values for risk measures, such as Value–at–Risk and Expected Shortfall, can increase as the level of

underlying asset correlation decreases. This is, at first sight, counterintuitive, as we would expect a higher correlation among asset values to imply more joint defaults. This is, indeed, the case; however, a higher correlation implies, at the same time, more joint "non–defaults". These zero–observations are able to outweigh the changes in the upper tail of the distribution.

The parameter ranges for which this phenomenon occurs may seem to hold only for few realistic portfolios. However, it has to be kept in mind that the strong correlations which we induced hold on the layer of latent variables and may translate into a much smaller *observed* correlation, according to Relationship (4.10). Furthermore, in case of highly–rated counterparties, default probabilities may be very small.

Whereas for Expected Shortfall this effect can be circumvented by choosing an appropriate design of the Monte-Carlo setup, this is, unfortunately, not so for the widely used Value–at–Risk, which systematically declines above certain levels of latent correlations. The extent to which this arises depends on the observed occurrence probabilities, the confidence level and the fat–tailedness of the distribution of the latent variables.
If the clustering of realizations at zero ("joint non–occurrences") is not in line with the true risk–generation mechanisms, risk capital may be severely underestimated. In this case, other statistical concepts of dependency should be considered for risk–capital calculation.

Chapter 5

Quantifying Operational Risk

5.1 The Operational Risk Challenge

The Basel Committee (Basel Committee on Banking Supervision, 2006) defines operational risk as "*the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events*". The broad range of loss events—including events such as bookkeeping errors, computer breakdowns, and terrorist attacks—calls for the separate modeling of seven event types and eight business lines. That is, in the standard Loss Distribution Approach (LDA), which is the most common approach to estimating operational–risk capital, a bank will have $i = 7 \times 8 = 56$ (possibly different) models for operational losses, L. This is sketched in Table 5.1.

		Business Lines		
		Corporate Finance	...	Retail Brokerage
Event Types	Internal Fraud	L_1	...	L_8
	⋮	⋮	⋱	⋮
	Execution, Delivery & Process Management	L_{49}	...	L_{56}

Table 5.1: The Business–Line/Event–Type Matrix

In the LDA, the losses of cell i are decomposed as

$$L_i = \sum_{n=1}^{N_i} X_{i,n} \, . \tag{5.1}$$

where N_i refers to the number of loss events and $X_{i,n}$ to the loss amounts for each event. As both components are themselves random, closed–form expressions for the compound distribution of L_i are available only for few parametric distributions. Typically, on has to resort to Monte Carlo simulation, the algorithm of Panjer (1981), the method of Heckman and Meyers (1983) or Fast Fourier Transforms (see, e.g., Robertson, 1992) in order to obtain the distribution of losses, L_i. A detailed account on this and other problems related to estimating risk capital for operational risk can be found in Panjer (2006).

Eventually, analyses have to condense into one single number, representing the aggregate operational–risk capital over all event types and business lines. This final quantity which one is interested in can be written as

$$\mathrm{VaR}_{.999}(L) = \mathrm{VaR}_{.999}\left(\sum_{i=1}^{56} L_i\right) \ . \tag{5.2}$$

Obviously, risk–capital estimates resulting from Equation (5.2) will be affected by dependencies among cells i and j. In other words, deriving risk capital along the lines of (5.2) requires an explicit modeling of dependencies.
In contrast to this, Basel II prescribes to calculate total risk capital (TRC) as

$$\mathrm{TRC} = \sum_{i=1}^{56} \mathrm{VaR}_{.999}(L_i) \ ; \tag{5.3}$$

only under certain qualifying conditions, banks may explicitly model dependencies. A natural question is thus: How will such an explicit modeling of dependencies, that is, a movement from Equation (5.3) to (5.2), affect risk–capital estimates? Put differently, can one be sure that working along the lines of Equation (5.2) will save regulatory capital? The answer to this is of paramount importance for assessing the conservativeness and the banks' incentives towards a more realistic modeling created by the Basel Committee.
VaR is a subadditive risk measure in the case of elliptical distributions, so that the standard Basel II calculation according to Equation (5.3) provides an upper bound and thus a worst–case scenario for aggregate risk. This illustrates why the idea of explicitly modeling dependencies according to (5.2) is very appealing: One intuitively expects a reduction in total risk–capital estimates. However, loss distributions are typically far from elliptical in the operational risk context. Banks may, thus, not be "rewarded" for a more realistic dependency modeling by a decrease in risk capital. We aim at assessing such diversification effects based on different measures of dependence.

A calculation of a minimum capital requirement according to (5.2) is influenced by

dependencies among frequencies, N, and severities, X. The Standard LDA[1] makes the following central assumptions:

1. N and $X_1, \ldots, X_N$ are independent,
2. $X_1, \ldots, X_N$ is a set of independent random variables,
3. $X_1, \ldots, X_N$ follow the same marginal distribution.

From this it follows that to include dependencies in the Standard LDA, only a correlation between frequencies can be considered without running into conceptual problems. Still, assuming correlated frequencies will translate into a correlation among aggregate losses, L_i. Frachot et al. (2004) derive that

$$\text{Corr}[L_1, L_2] \leq \text{Corr}[N_1, N_2] \ , \tag{5.4}$$

and the difference between the two will be the higher, the higher the severities of loss events (due a domination of severity independence over frequency dependence). Thus, even for highly correlated frequencies, aggregate losses may exhibit only small correlations.

Copulas can, in principle, be applied to the continuous severity distributions as well as to the discrete counting variables; in the latter case, they are not necessarily unique. Analyses and applications of copulas to the numbers of events can be found, for example, in Pfeifer and Nešlehová (2004) and Chavez-Demoulin et al. (2006). In the following, we do not distinguish between the source of dependence; rather, we aim at directly capturing the dependence structure among aggregate losses, L_i and L_j.

As mentioned by Chavez-Demoulin et al. (2006), the lack of reliable operational loss data has impeded research on a possible reduction of the calculated risk capital by explicitly considering dependencies. Analyses often concentrate on modeling the marginal distributions only. The few studies based on real–world data which treat this dependence issue, such as Chapelle et al. (2004) or Aue and Kalkbrener (2006), find that the prescribed assumption of perfect correlation indeed leads to an overestimation of risk capital. Using mixture models, Reshetar (2008) concludes that the prescribed assumption of perfect correlation indeed leads to an overestimation of risk capital. Cope and Antonini (2008/2009), analyzing data from the ORX data base, consider estimation of correlations as well as tail dependence; they find only slight evidence of tail dependence. While they find diversification benefits for high quantiles, this effect is found to be potentially reversed for lower confidence levels.

[1]The Standard LDA without consideration of dependencies is described and analyzed in detailed, by Frachot et al. (2001) and Frachot et al. (2003).

5.2 Estimation and Effects of Dependencies

Drawing on real–world operational risk data from the DIPO database[2], our aim is to assess the relevance of the potential superadditivity of VaR. The quantity we thus consider is

$$D = \frac{\mathrm{VaR}_{.999}(L_i + L_j) - (\mathrm{VaR}_{.999}(L_i) + \mathrm{VaR}_{.999}(L_j))}{(\mathrm{VaR}_{.999}(L_i) + \mathrm{VaR}_{.999}(L_j))} \tag{5.5}$$

as measure of diversification; positive values imply that an explicit modeling of dependencies leads to an increase in risk–capital estimates compared to the simple summing–up of VaRs.

5.2.1 Modeling Dependencies

The first and most obvious attempt at capturing the dependence structure among aggregate losses consists of estimating linear correlation. The results are presented in Table 5.2. The upper part of the matrix contains correlation estimates using the full sample, while the lower part restricts attention to 2/3 of the data.

Without considering the effect different samples may have on estimation (i.e., sticking to the upper part of Table 5.2), we observe that correlation may vary substantially, depending on the event types considered. For example, event type combination 2/5 (External Fraud/Damage to Physical Assets) exhibits a quite small correlation, while combination 3/4 (Employment Practices & Workplace Safety/Clients, Products & Business Practices) is characterized by a substantial correlation of $\rho = 0.528$. This first finding tells us to differentiate between event types when drawing conclusions about dependence structures in operational risk data.

As they represent the range of possible dependence structures, we will continue to restrict attention to event type combinations 2/5 and 3/4 in the following. However, this does not imply that the conclusions thereby obtained are genuine to these combinations; they rather represent a spectrum of different behavior which may occur.
Next, we consider the change in correlation estimates resulting from a reduction of the sample, i.e., we compare the upper and lower parts of Table 5.2. We observe that correlation estimates are unstable, which is a well–known deficiency of linear correlation. This is illustrated in Figure 5.1, which shows scatter plots of aggregate losses for both event type combinations and sample sized. For example, the reduction of correlation observed for event type combination 2/5 corresponds to two extreme observations (marked by circles) which drop out of the sample in the course of sample

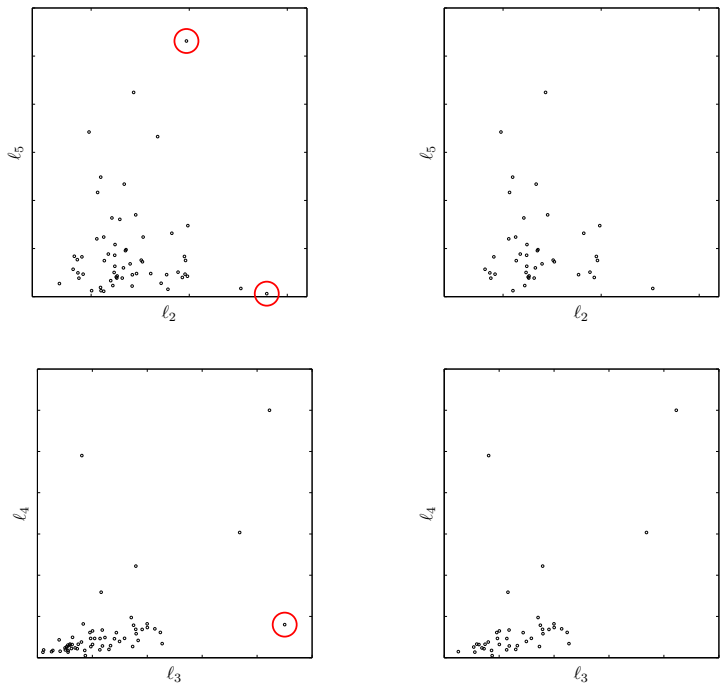
[2] `http://www.dipo-operationalrisk.it`

	ET 1	ET 2	ET 3	ET 4	ET 5	ET 6	ET 7
ET 1	1	0.045	0.221	0.046	0.010	0.085	0.223
ET 2	0.062	1	0.162	0.161	**0.026**	-0.066	0.052
ET 3	0.126	0.283	1	**0.528**	0.295	0.182	0.443
ET 4	-0.028	0.242	**0.588**	1	0.143	0.063	0.201
ET 5	-0.012	**-0.114**	0.290	0.177	1	-0.046	-0.050
ET 6	0.030	-0.039	0.151	0.012	-0.053	1	0.065
ET 7	0.155	0.063	0.365	0.202	-0.104	0.045	1

Upper part: 01/03–12/07, lower part: 01/03–04/06

Table 5.2: Estimated Linear Correlation Coefficients

reduction. Similarly, the drop in correlation for event type combination 3/4 coincides with the elimination of an extreme observation.



01/03–12/07 01/03–04/06

Top: event type combination 2/5; bottom: event type combination 3/4.
Left: full sample; right: 2/3 of sample.

Figure 5.1: Scatterplots of Aggregate Losses

	ET 1	ET 2	ET 3	ET 4	ET 5	ET 6	ET 7
ET 1	1	0.135	0.248	0.214	0.142	0.105	0.190
ET 2	0.097	1	0.148	0.211	**0.005**	-0.024	0.111
ET 3	0.064	0.192	1	**0.510**	0.260	0.207	0.183
ET 4	0.049	0.218	**0.451**	1	0.307	0.313	0.072
ET 5	0.046	**-0.005**	0.162	0.356	1	0.052	-0.033
ET 6	0.041	0.097	0.151	0.280	0.051	1	0.172
ET 7	0.133	0.082	0.090	-0.044	-0.139	0.139	1

Upper part: 01/03–12/07, lower part: 01/03–04/06

Table 5.3: Estimated Kendall Rank Correlation Coefficients

	ET 1	ET 2	ET 3	ET 4	ET 5	ET 6	ET 7
ET 1	1	0.197	0.343	0.312	0.223	0.152	0.293
ET 2	0.157	1	0.202	0.293	**0.001**	-0.029	0.153
ET 3	0.057	0.276	1	**0.695**	0.385	0.320	0.257
ET 4	0.071	0.301	**0.633**	1	0.433	0.451	0.109
ET 5	0.096	**-0.002**	0.258	0.494	1	0.079	-0.043
ET 6	0.062	0.146	0.236	0.411	0.088	1	0.272
ET 7	0.191	0.112	0.125	-0.045	-0.202	0.226	1

Upper part: 01/03–12/07, lower part: 01/03–04/06

Table 5.4: Estimated Spearman Rank Correlation Coefficients

Knowing that correlation may lead to very volatile results, an obvious alternative is thus given by rank correlations. However, as is revealed by Tables 5.3 and 5.4, it is not only the loss magnitudes which change when reducing the sample size, but also their ranks, so that the estimated values of Kendall's τ and Spearman's ρ are unstable as well.

Furthermore, just as Pearson's correlation, rank correlation boil down dependencies into one single number and are thus not sufficiently flexible. We thus conclude that, for our purposes, linear correlation as well as rank correlation is an inappropriate measure of dependence.

Copulas

In order to analyze effects of different dependency structures on VaR, we fit several copulas via the IFM method to the data on an event–type level.
We do not estimate any parametric marginal distributions, but rather use the empirical cdf as inputs to the likelihood function.[3] The rationale behind this procedure is that we do not want to impose restrictions which could have an influence on the estimation results for the dependence structure.

	$\widehat{\Theta}$	$\widehat{\lambda}_U$	ℓ	AIC	BIC	p–value $\mathcal{A}_2$	p–value $\mathcal{A}_4$	d
	01/03–12/07							
Ga	-0.015	0.000	-0.005	1.989	8.178	0.640	0.619	0.545
t	3.86	0.095	-0.945	0.110	6.299	0.798	0.606	0.804
Gu	1.015	0.020	-0.016	1.969	8.157	0.790	0.769	0.690
GuSu	1.000	0.000	0.000	2.000	8.189	0.697	0.743	0.599
Cl	0.023	0.000	-0.009	1.982	8.171	0.792	0.775	0.642
ClS	0.002	0.000	-0.000	2.000	8.189	0.692	0.628	0.604
	01/03–04/06							
Ga	-0.014	0.000	-0.003	1.994	8.183	0.7053	0.599	1.342
t	7.51	0.027	-0.092	1.817	8.005	0.7592	0.645	1.446
Gu	1.00	0.000	0.000	2.001	8.189	0.6254	0.613	1.449
GuS	1.01	0.000	-0.001	1.998	8.187	0.6943	0.670	1.476
Cl	0.081	0.000	-0.068	1.864	8.053	0.8581	0.849	1.901
ClS	0.000	0.000	0.000	2.000	8.189	0.6583	0.648	1.436

Table 5.5: Maximum–Likelihood Estimation Results for Different Parametric Copulas, Event Type Combination 2/5.

The estimation results for event type combination 2/5 can be found in Table 5.5. While the upper part shows estimations results from the full sample, the lower part is again based on 2/3 of the sample. The first column contains the parameter estimates obtained by the semiparametric maximum likelihood procedure. For example, using a Gaussian copula, the estimated correlation is $\widehat{\rho} = -0.015$ in the full sample, changing to $\widehat{\rho} = -0.014$ in the smaller sample. This is a stronger and, above all, a more stable correlation estimate than the Pearson one. The second column contains the upper tail

[3] Again, we use the rescaled version of Genest et al. (1995).

dependence coefficient which is implied by the estimation results; it is derived using the relationships from Table 2.3. Among the three copulas being able to capture upper tail dependence, only the Student-t leads to considerable tail dependence; for the Gumbel and even more for the Clayton survival, we find only negligible tail dependence.
The remaining columns aim at assessing the goodness of fit. Both AIC and BIC point at the Student-t copula in the full as well as in the small sample. However, while this result is clear–cut in the full sample, the AIC and BIC values are quite close in the small sample. The next two columns present the p–values from the two goodness–of–fit tests found to perform considerably well by Berg (2009): The $\mathcal{A}_2$ test of Genest and Rémillard (2008) and the $\mathcal{A}_4$ test which can be found, for example, in Genest et al. (2006).[4] However, none of the p–values derived from a bootstrap procedure allows the rejection of a copula. The last column contains the sum of squared distances between empirical and theoretically implied extremal dependence functions as presented in Section 2.3.3.4. The values are close, with exception of the Student–t copula for the full and the Clayton copula for the small sample. Due to the small differences in results, we do not take any decision as to which copula should be preferred based on this criterion here.

Table 5.6 shows the corresponding results for event type combination 3/4. Here, we do find evidence for tail dependence: AIC and BIC choose for both samples a copula leading to substantial tail dependence (in the full sample, the Student–t with $\widehat{\lambda} = 0.40$, in the small sample the Gumbel copula with $\widehat{\lambda} = 0.56$). The preferred copula now depends on the sample; furthemore, the values of the information criteria are quite close. With exception of the Clayton survival copula in the full sample, a rejection of any copula based on goodness–of–fit tests is also not possible here.
An illustration of the fit of the different copulas is provided by Figure 5.2, which shows contour plots of the empirical and the fitted copulas. The left part is based on the full sample, while the right part considers only 2/3 of it. We see that while for event type combination 2/5 (upper part), all parametric copula contours are close, they vary for event type combination 3/4 (lower part). This corresponds to the fact that all fitted copulas exhibit absence of tail dependence for combination 2/5, while those allowing for it lead to substantial tail dependence for combination 3/4.

The general challenge of parametric copula estimation in small samples is illustrated by the kinks of the empirical copulas, caused by single observations. We see that reducing the sample leads to noticeable differences in the empirical copula, so that stability remains an issue.

[4]We use the same numbers of replications as Berg (2009), that is, 10,000 replications for the parametric bootstrap and an additional number of 2,500 replications in the double bootstrap.

	$\widehat{\Theta}$	$\widehat{\lambda}_U$	ℓ	AIC	BIC	p–value $\mathcal{A}_2$	p–value $\mathcal{A}_4$	d
	01/03–12/07							
Ga	0.719	0.000	-19.489	-36.977	-30.789	0.912	0.920	2.745
t	6.618	0.402	-19.971	-37.941	-31.752	0.961	0.935	3.319
Gu	1.954	0.574	-18.386	-34.772	-28.583	0.285	0.499	3.945
GuS	1.986	0.000	-19.443	-36.885	-30.696	0.859	0.828	5.667
Cl	1.522	0.000	-17.065	-32.129	-25.940	0.103	0.228	3.442
ClS	1.398	0.609	-15.15	-28.300	-22.111	0.010	0.048	8.859
	01/03–04/06							
Ga	0.682	0.000	-10.593	-19.1864	-12.998	0.8452	0.9071	4.574
t	5.865	0.390	-11.023	-20.0453	-13.857	0.9181	0.8362	5.245
Gu	1.910	0.562	-11.315	-20.6297	-14.441	0.6134	0.7293	5.618
GuS	1.795	0.000	-9.630	-17.2595	-11.071	0.5594	0.8561	4.594
Cl	1.193	0.000	-8.341	-14.6819	-8.493	0.0909	0.3117	4.590
ClS	1.494	0.629	-10.017	-18.0344	-11.846	0.1978	0.3357	8.205

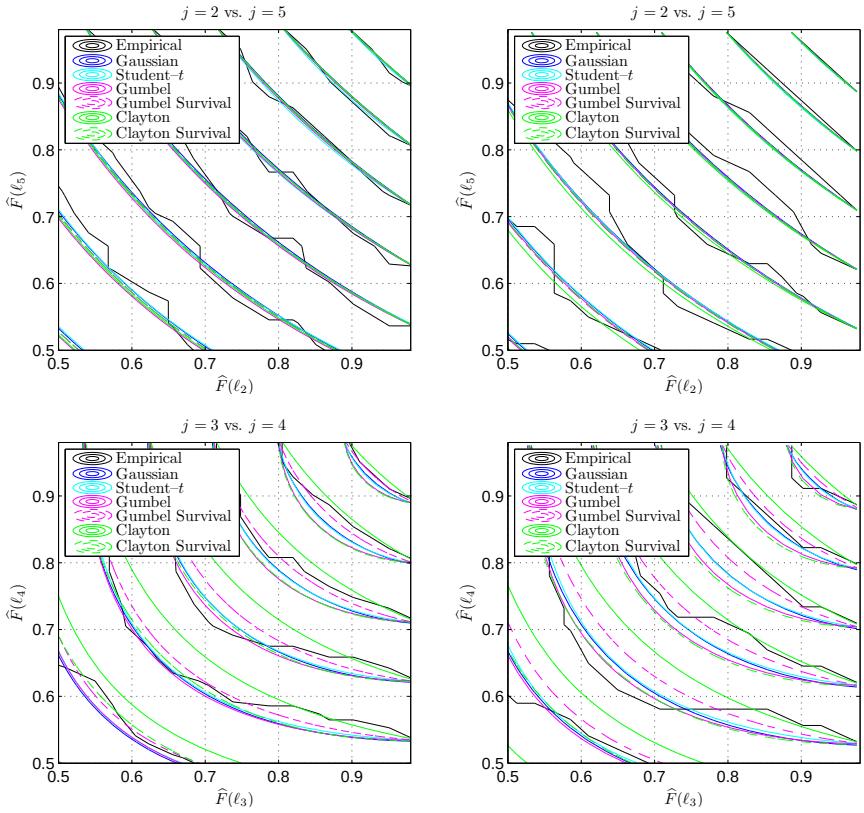
Table 5.6: Maximum–Likelihood Estimation Results for Different Parametric Copulas, Event Type Combination 3/4.

Nonparametric Tail Dependence

Figures 5.3 and 5.4 show the results of nonparametric estimation. For event type combination 2/5, we do not find any evidence of tail dependence nor quantile dependence: The estimators of λ_U, $\widehat{\lambda}_U(t)$ and $\widehat{\chi}(t)$ do not converge to a positive level as t increases. At the same time, $\widehat{\bar{\chi}}(t)$ is close to zero over the entire range of t, so that quantile dependence does not seem to be an issue either. For event type combination 3/4, the results are different. Although $\widehat{\lambda}_U(t)$ and $\widehat{\chi}(t)$ drop towards zero for $t > 0.75$, $\widehat{\bar{\chi}}(t)$ is clearly positive over a wide range of t, hinting at quantile dependence.

In Figures 5.3 and 5.4, the theoretically implied extremal dependence functions derived from Maximum Likelihood estimation are compared to the nonparametric estimates. We see that for event type combination 2/5, they are very close, and a decision towards a parametric copula cannot be taken based on their shapes. In contrast to this, for event type combination 3/4 the Clayton copula implies extremal dependence functions closest to the empirically observed ones.

A general remark has to be made with respect to high levels of t. The higher t, the less data points are used for estimation. That is, as we move towards the levels of t which

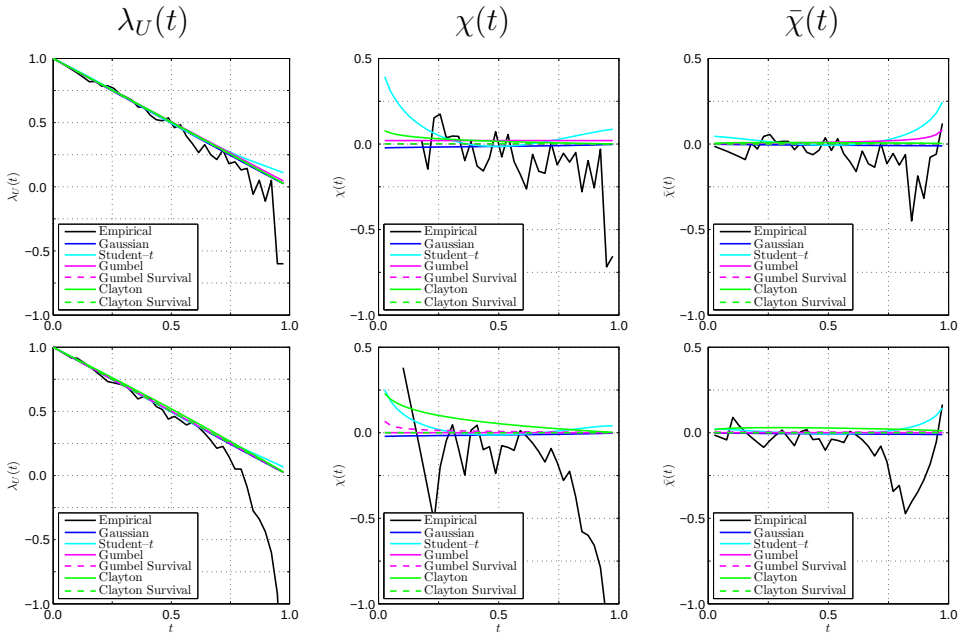


Top: event type combination 2/5; bottom: event type combination 3/4.
Left: full sample; right: 2/3 of sample.

Figure 5.2: Contour Plots of Empirical and Fitted Copulas

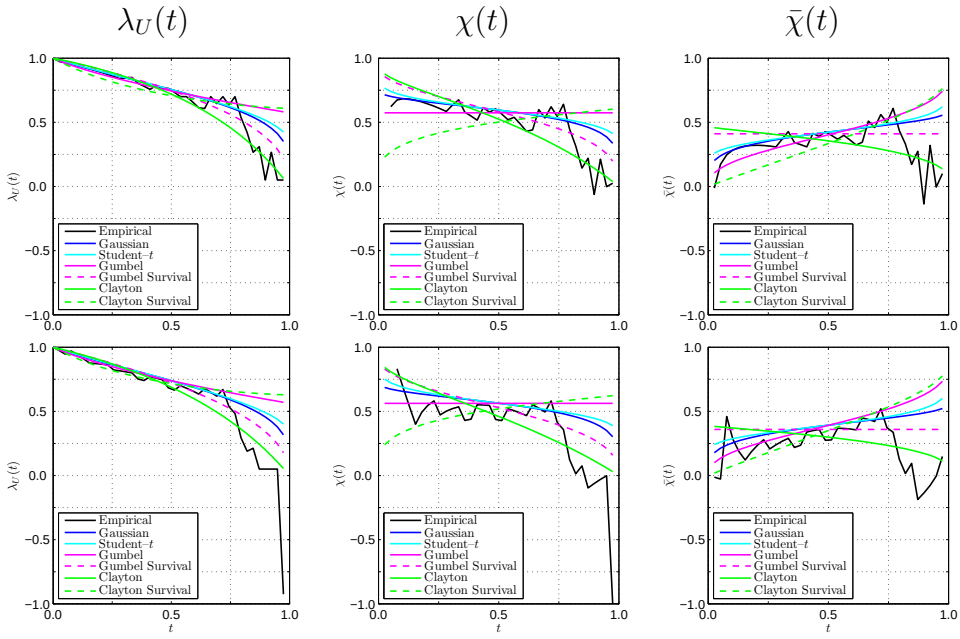
are relevant for detecting tail dependence, our estimation results become less and less reliable. Inferring absence of tail dependence from values of $\widehat{\lambda}_U(t)$ and $\widehat{\chi}(t)$ for $t \geq 0.9$, which are based on very few data points, might therefore be a misleading strategy.

Instead, we advocate the joint use of all available information from the different approaches to dependency modeling presented. The results from the parametric copula fitting, suggesting a tendency towards tail dependence for event type combination 3/4 but absence of it for combination 2/5, can be recovered using nonparametric methods. We again find different patterns for the two combinations, confirming that there is no one–size–fits–all statement about dependency structures in losses from operational risk.



Top: full sample; bottom: 2/3 of sample.

Figure 5.3: Nonparametric and Implied Extremal Dependence, Event Type Combination 2/5



Top: full sample; bottom: 2/3 of sample.

Figure 5.4: Nonparametric and Implied Extremal Dependence, Event Type Combination 3/4

5.2.2 Effects on Risk Capital

Range of Risk–Capital Estimates

Under the standard LDA approach, banks will sum up their single risk–capital estimates according to Equation (5.3). The question which we aim at answering is: Will a more realistic dependency modeling lead to a decrease in aggregate risk capital, or are the loss distributions' properties such that superadditivity of Value–at–Risk may become relevant?

To this end, we estimate 250 $\text{VaR}_{99.9}$ figures per dependency model and event type combination, using different numbers of replications. For each event type combination, we use the copula parameter values obtained from Maximum Likelihood estimation. In order to be able to simulate losses without distorting results by the use of different marginal distributions, a lognormal distribution is fitted to each aggregate loss distributions and used to derive risk–capital estimates via simulation.

The quantity we eventually consider is

$$\frac{\text{VaR}_{.999}(L_i + L_j) - (\text{VaR}_{.999}(L_i) + \text{VaR}_{.999}(L_j))}{(\text{VaR}_{.999}(L_i) + \text{VaR}_{.999}(L_j))} \,. \tag{5.6}$$

That is, we evaluate the percentage difference in risk capital achieved by an explicit modeling of dependencies. A positive value of this quantity will thus imply that risk–capital estimates increase as compared to the Basel II standard approach. For the dependence scenarios used as alternatives to the comonotonicity assumption, we consider two extreme cases: Firstly, we assume the Gaussian copula for all event type combinations; secondly, we use a "worst–case" copula, i.e., that copula yielding the highest tail dependence coefficient for the respective event type combination.

Figure 5.5 shows results using the Gaussian copula. For $B_{rc} = 10{,}000$ replications, the differences among event type combinations can already be detected. For combination 2/5, which was characterized by a negative correlation estimate, risk–capital estimates always decrease, and a reduction in risk–capital of up to almost 30% is possible. However, for combination 3/4, a use of the Gaussian copula may lead to an increase in risk capital, caused by the correlation estimate of about $\widehat{\rho} \approx 0.7$. Here, it may occur that the risk–capital estimate is almost 30% higher than the Basel standard calculation. Still, the interquartile range is entirely below zero, so that one would observe a decrease in most cases.

Increasing the number of replications leads to a narrowing of the boxplots, and thus to less drastic changes in risk–capital estimates. For $B_{rc} = 50{,}000$ replications, the

possible range for combination 2/5 lies between -19% and -25%; for $B_{rc} = 100{,}000$, this range becomes even smaller. Considering event type combination 3/4, we see that, correspondingly, the higher the number of replications, the smaller the number of increases in risk–capital estimates. Using $B_{rc} = 100{,}000$ replications, the maximum increase in risk–capitel is $+3\%$.

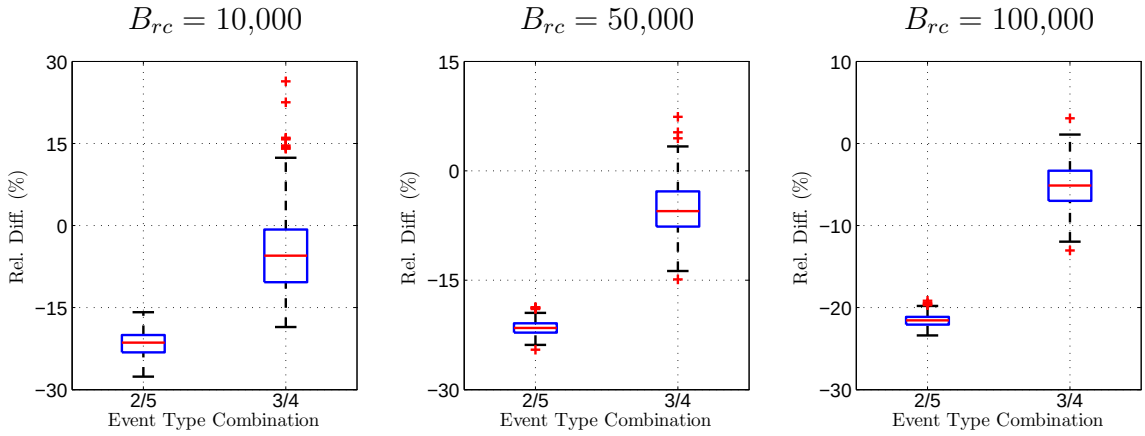


Figure 5.5: Range of Risk–Capital Estimates: Gaussian Copula

Figure 5.6 shows the result for the worst–case copula, which is the Student–t (implying a tail dependence coefficient of $\lambda_U = 0.095$) for combination 2/5 and the Clayton survival ($\lambda_U = 0.629$) for combination 3/4. The difference in the copula choice as compared to Figure 5.7 leads to an upward shift of the boxplots: The presence of tail dependence leads to more cases of risk–capital increases. For the case of low tail dependence (2/5), however, risk–capital decreases for all numbers of replications used; for the case of high tail dependence (3/4), increases in risk–capital may be up to 6% for $B_{rc} = 100{,}000$ replications. Now, even for $B_{rc} = 100{,}000$ replications, the interquartile range for combination 3/5 includes zero.

Obviously, there are two different effects which may cause increases in risk–capital estimates. Firstly, the superadditivity property of VaR may be the reason; secondly, the computational setup, that is, the number of replications in the Monte Carlo simulation has an influence on estimates. It is therefore desirable to be able to disentangle these two effects.

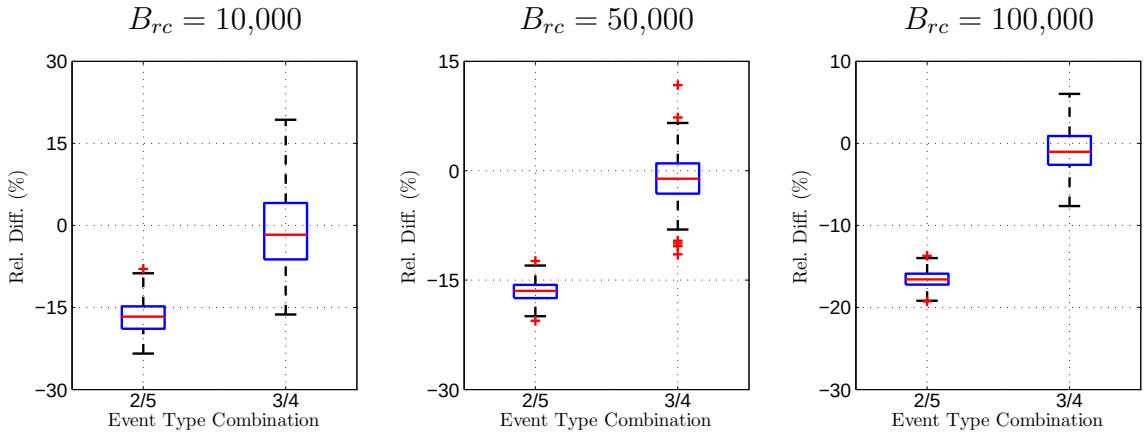


Figure 5.6: Range of Risk–Capital Estimates: Worst–Case Copula

Bounds on Risk–Capital Estimates

In the following, we analyze bounds on risk–capital estimates derived under different assumptions with respect to the copula. As described in Section 2.4.3, such bounds are tighter, the more restrictive the assumptions on the copula bounds. We consider three different cases of increasing restrictiveness:

1. $C_0 = C_1 = C_\ell$: We do not use any restriction on the dependence structure and thus use the lower Fréchet bound, C_ℓ.

2. $C_0 = C_1 = u_i u_j$: We assume that $C \geq u_i u_j$, that is, we have positive quadrant dependence (PQD).

3. $C_0 = C_{\text{CS}}^{\widehat{\theta}_{i,j}}, \widehat{C}_1 = C_C^{\widehat{\theta}_{i,j}}$: We take the Clayton Survival copula as lower bound, using the parameter values estimated for the DIPO data. For the survival copula of C_1, we accordingly assume the Clayton copula with respective parameter values.

Figures 5.7 and 5.8 show once again the boxplots as presented in the previous section, but now the theoretical bounds on risk–capital changes are included. We observe the increasing tightness of the bounds caused by stronger dependence assumptions. For $B_{rc} = 10{,}000$, a large part of the boxplots exceeds the theoretical bounds; as the number of replications increases, the larger the part of estimates lying inside them. For event type combination 3/4 and $B_{rc} = 100{,}000$ replications, only the most restrictive bounds are exceeded by capital estimates.

These bounds may, therefore, be used for smaller numbers of replications in order to detect whether a change in risk–capital is caused by the distributional properties or

by computational issue. However, for none of the two event type combinations do the bounds lie entirely above or below zero, so that we cannot say whether an increase or decrease is theoretically impossible.

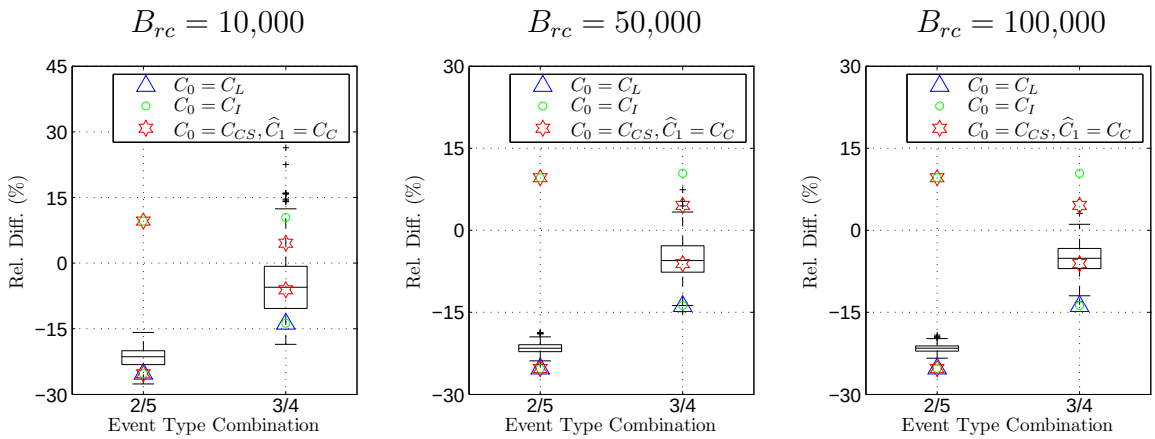


Figure 5.7: Bounds on Risk Capital: Gaussian Copula

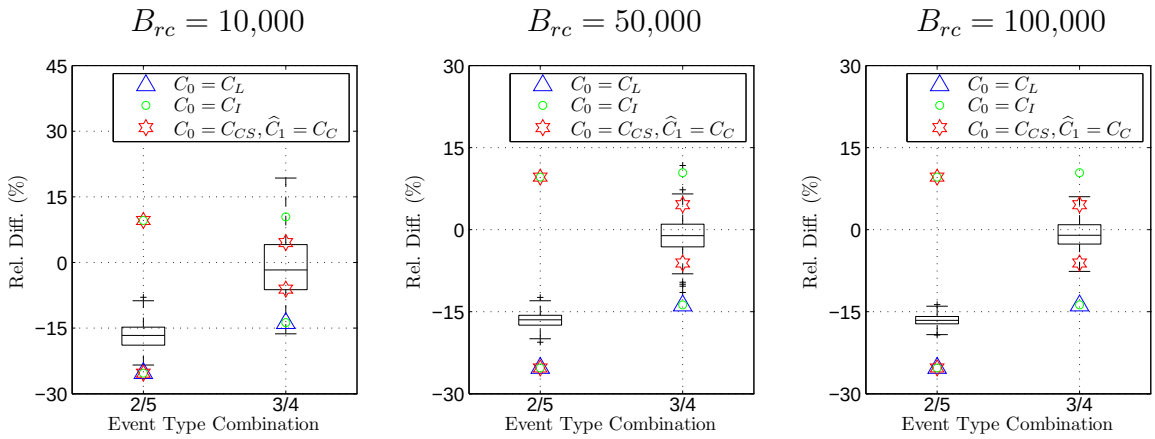


Figure 5.8: Bounds on Risk Capital: Worst–Case Copula

5.3 Chapter Summary

It is known that diversification effects depend—among other factors—on the presence and nature of dependence among the underlying variables. For operational risk, the question whether an explicit modeling of dependencies may lower regulatory capital charges has not been answered yet, which is mainly due to data scarcity. Based on real–world data, we assess dependence using different measures and estimation procedures, as well as their effects on risk–capital estimates.

We find that the change from a Gaussian copula to one implying substantial tail dependence may in fact lead to an upward shift of risk–capital; thus, the question of whether VaR will increase or decrease crucially depends on the presence of extremal (tail/quantile) dependence and the ellipticity of the multivariate distribution. We obtain different results for different event–type combinations, so that a general answer as to whether there exists tail dependence or not should not be given. Therefore, the estimation of extremal dependence for the data at hand is of paramount importance for an assessment of the relevance of this effect. Neglecting this dependency concept and restricting oneself to linear correlation estimation may lead to a substantial underestimation of risk.

On the other hand, inference is complicated by the small sample sizes one typically has to deal with in operational risk settings. That is, copula estimation and goodness–of–fit methods as well as nonparametric methods are to be taken with care; and we recommend to jointly consider these different methods of inference in order to derive a consistent, complete picture of dependencies.

Chapter 6

Summary and Conclusions

Effects of dependencies are a crucial element to risk quantification in an aggregate setting, for example, when several risk types or different lines of business need to be considered jointly. Diversification among the components of a portfolio has become a standard notion, and it is common wisdom to avoid putting all one's eggs into one basket. However, such ideas are inherently linked to elliptical distributions. For those, linear correlation is a meaningful measure of dependence, and the popular VaR risk measure can be shown to be subadditive, so that the sum of single risks provides a worst–case scenario.

However, risk management typically has to deal with non–elliptical distributions. Each of the risk types poses specific challenges to risk quantification under consideration of dependencies. We have considered three cases of such special challenges, where standard methods may prove to be misleading or should at least be questioned. For market risks, besides the well–known univariate stylized facts, correlations are known to increase in times of large negative movements. These empirically observed conditional correlations cannot be united with the assumption of a constant Gaussian dependency structure; this in turn implies that portfolio selection based on elliptical distributions may lead to an overestimation of diversification benefits in times of crises.

Credit risks prove to be especially difficult to quantify in situations of low default probabilities, that is, for highly rated counterparties or very short–term exposures. A quantification of default risk needs to be based on simulation—proving to be a challenge in itself due to the substantial computational burden connected to rare–event simulation. We suggest to solve this technical problem by relying on quasi–random sequences. In the frame of a simulation study, we have shown that—without any adding up of risk capital estimates—linear correlation may have very counterintuitive results on portfolio risk. That is, an increase in latent (asset) correlations of the components

of the portfolio can lead to a decrease in risk, as measured by the VaR of the aggregate loss distribution. This effect is due to the increase in joint "non–defaults" caused by a higher correlation and should be taken into account when simulating rare events.

For operational risks, the challenges to quantification are manifold due to the scarcity of data, the high confidence level required by the Basel Committee and the dimensionality of the problem, among other factors. Based on real–world data, we have shown that the well–known lack of subadditivity of the VaR risk measure may in fact become relevant. This was shown under the assumption of a "worst–case" copula, that is, we chose the highest tail dependence obtained from estimation. At the same time, a part of potential increases are merely due to computational issues, so that theoretical bounds on VaR may help to assess whether a certain change in risk capital is realistic.

For all these cases, the necessary first step of analyses was the estimation of the dependency structure. Linear correlation was shown to neglect nonlinear effects and joint extremes, and should therefore always be handled with care and joined by other measures of dependency. The concepts of copulas and tail dependence may be used for flexibly analyzing dependencies in the tails of loss distributions. However, both approaches—which are, in fact, linked to each other—suffer from substantial limitations in practical use. Distinguishing between different copulas may be difficult in small samples and for moderate levels of dependence. As for nonparametric estimation of tail dependence, the size of the sample used for estimation decreases as one moves into the tails of the distributions. In other words, estimation of extremal dependence is most difficult when it is needed the most.

As these issues cannot be simply resolved given a certain dependence structure and sample size, they must be taken into account. This can be done by considering as many measures of dependence and goodness–of–fit tests as possible. In this spirit, we have suggested a path for a new criterion for the goodness–of–fit of parametric copulas. This criterion is based on a comparison of nonparametric tail dependence estimators and parametric copulas. It provides an additional instrument for assessing the goodness–of–fit of copulas. We do, however, not recommend to rely on this (or any other) criterion exclusively, but to gain a picture as comprehensive as possible through different estimators of dependence.

Bibliography

Agmon, T., 1972. The Relations Among Equity Markets: A Study of Share Price Co–Movements in the United States, United Kingdom, Germany and Japan. Journal of Finance 27 (4), 839–55.

Aït-Sahalia, Y., Brandt, M. W., 2001. Variable Selection for Portfolio Choice. Journal of Finance 56, 1297–1355.

Albrecher, H., Hartinger, J., Tichy, R. F., 2003. Multivariate approximation methods for the pricing of catastrophe-linked bonds. International Series of Numerical Mathematics 145, 21–39.

Anderson, T. W., Darling, D. A., 1952. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. Annals of Mathematical Statistics 23, 193–212.

Ang, A., Bekaert, C., 2002. International asset correlation with regime shifts. Review of Financial Studies 15, 1137–1187.

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. Journal of Financial Economics 63, 443–494.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent Measures of Risk. Mathematical Finance 9, 203–228.

Aue, F., Kalkbrener, M., 2006. LDA at Work: Deutsche Bank's Approach to Quantifying Operational Risk. Journal of Operational Risk 1 (4), 49–93.

Babu, G. J., Rao, C. R., 2004. Goodness–of–fit tests when parameters are estimated. Sankhyā 66 (1), 63–74.

Balkema, G., Embrechts, P., 2007. High Risk Scenarios and Extremes: A Geometric Approach. European Mathematical Society.

Balyeat, R. B., Muthuswamy, J., 2009. The Correlation Structure of Unexpected Returns in U.S. Equities. Financial Review 44, 269–290.

Barbe, P., Genest, C., Ghoudi, K., Rémillard, B., 1996. On Kendall's Process. Journal of Multivariate Analysis 58, 197–229.

Barlow, R., Proschan, F., 1981. Mathematical Theory of Reliability. To Begin With.

Basel Committee on Banking Supervision, June 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework. Technical Report, Bank for International Settlements, Comprehensive Version.

Bawa, V. S., 1975. Optimal rules for ordering uncertain prospects. Journal of Financial Economics 2, 95–121.

Bawa, V. S., 1978. Safety first, stochastic dominance and optimal portfolio choice. Journal of Financial and Quantitative Analysis 13 (2), 255–271.

Berg, D., 2009. Copula goodness-of-fit testing: An overview and power comparison. European Journal of Finance 15, 675–701.

Berg, D., Bakken, H., 2005. Goodness–of–fit test for copulae based on the probability integral transform. Statistical Research Report 10, University of Oslo.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business and Economic Statistics 19 (4), 465–474.

Blattberg, R. C., Gonedes, N., 1974. A comparison of the stable and student distributions as statistical methods for stock prices. Journal of Business 47, 244–280.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307–327.

Bortot, P., Tawn, J. A., 1998. Models for the extremes of Markov chains. Biometrika 85, 851–867.

Boyer, B. H., Gibson, M. S., Loretan, M., 1999. Pitfalls in tests for changes in correlations. International Finance Discussion Paper 597, Board of Governors of the Federal Reserve System.

Breymann, W., Dias, A., Embrechts, P., 2003. Dependence structures for multivariate high-frequency data in finance. Quantitative Finance 1, 1–14.

Bruun, J. T., Tawn, J. A., 1998. Comparison of approaches for estimating the probability of coastal flooding. Journal of the Royal Statistical Society: Series C 47 (3), 405–423.

Caflisch, R., Morokoff, W., Owen, A., 1997. Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension. Journal of Computational Finance 1, 27–46.

Campbell, R. A. J., Koedijk, K. G., Kofman, P., 2002. Increased correlation in bear markets. Financial Analysts Journal 58, 87–94.

Capéraà, P., Fougères, A.-L., 2000. Estimation of a bivariate extreme value distribution. Extremes 3, 311–329.

Capéraà, P., Fougères, A.-L., Genest, C., 1997. A nonparametric estimation procedure for bivariate extreme value copulas. Biometrika 84, 567–577.

Chapelle, A., Crama, Y., Hübner, G., Peters, J.-P., 2004. Basel II and Operational Risk: Implications for risk measurement and management in the financial sector. Working Paper 51, National Bank of Belgium.

Chavez-Demoulin, V., Embrechts, P., Nešlehová, J., 2006. Quantitative Models for Operational Risk: Extremes, Dependence and Aggregation. Journal of Banking and Finance 30, 2635–2658.

Chekhlov, A., Uryasev, S., Zabarankin, M., 2005. Drawdown measure in portfolio optimization. International Journal of Theoretical and Applied Finance 8 (1), 13–58.

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. Wiley.

Chua, J., Sick, G., Woodward, R., 1990. Diversifying with gold stocks. Financial Analysts Journal 46, 76–79.

Clayton, D. G., 1978. A model for association in bivariate life tables and its application in epidemicological studies of familial tendency in chronic disease incidence. Biometrika 65, 141–151.

Coles, S. G., Heffernan, J. E., Tawn, J. A., 1999. Dependence measures for extreme value analyses. Extremes 2, 339–365.

Coles, S. G., Tawn, J. A., 1994. Statistical Methods for Multivariate Extremes: An Application to Structural Design. Applied Statistics 43 (1), 1–48.

Cook, D., Johnson, M. E., 1981. A family of distributions for modeling non–elliptically symmetric multivariate data. Journal of the Royal Statistical Society. Series B 43 (2), 210–218.

Cope, E., Antonini, G., 2008/2009. Observed correlations and dependencies among operational losses in the ORX consortium database. Journal of Operational Risk 3 (4), 47–74.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C., 2001. Introduction to Algorithms. MIT Press.

Corsetti, G., Pericoli, M., Sbracia, M., 2005. 'Some contagion, some interdependence': More pitfalls in tests of financial contagion. Journal of International Money and Finance 24 (8), 1177–1199.

D'Agostino, R. B., Stephens, M. A. (Eds.), 1986. Goodness of Fit Techniques. Marcel Dekker.

Deheuvels, P., 1979. La function de dépendance empiriques et ses propriétés. Un test non paramétrique d'indépendance. Académie Royale de Belgique—Bulletin de la Classe des Sciences 65, 274–292.

Deheuvels, P., 1981. A Nonparametric Test for Independence. Publications de l'Institut de Statistique de l'Université de Paris 26.

Dias, A., Embrechts, P., 2004. Dynamic copula models for multivariate high–frequency data in finance. Mimeo, Department of Mathematics, New University of Lisbon.

Dobric, J., Schmid, F., 2005. Testing Goodness of Fit for Parametric Families of Copulas—Application to Financial Data. Communications in Statistics—Simulation and Computation 35, 1053–1068.

Dobric, J., Schmid, F., 2007. A goodness of fit test for copulas based on Rosenblatt's transformation. Computational Statistics & Data Analysis 51 (9), 4633–4642.

Dornbusch, R., Park, Y.-C., Claessens, S., 2001. Contagion: How It Spreads and How It Can be Stopped. In: International Financial Contagion. Kluwer Academic.

Duchateau, L., Janssen, P., 2008. The Frailty Model. Springer.

Dupuis, D. J., Jones, B. L., October 2006. Multivariate Extreme Value Theory and Its Usefulness in Understanding Risk. North American Actuarial Journal 10 (4), 1–27.

Efron, B., Tibshirani, R., 1986. Bootstrap method for standard errors, confidence intervals and other measures of statistical accuracy. Statistical Science 1 (1), 54–75.

Embrechts, P., Höing, A., Juri, A., 2003. Using Copulae to bound the Value-at-Risk for functions of dependent risks. Finance & Stochastics 7 (2), 145–167.

Embrechts, p., Klüppelberg, C., Mikosch, T., 2007. Modelling Extremal Events. Springer.

Embrechts, P., Lambrigger, D. D., Wüthrich, M. V., 2009. Multivariate extremes and the aggregation of dependent risks: examples and counter–examples. Extremes 12, 107–127.

Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M. (Ed.), Risk Management: Value at Risk and Beyond. Cambridge University Press.

Embrechts, P., Puccetti, G., 2010. Bounds for the sum of dependent risks having overlapping marginals. Journal of Multivariate Analysis 101, 177–190.

Enterprise Risk Management Committee, 2003. Overview of Enterprise Risk Management. Technical report, Casualty Actuarial Society, `http://www.casact.org/research/erm/overview.pdf`.

Esary, J. D., Proschan, F., 1972. Relationships Among Some Concepts of Bivariate Dependence. The Annals of Mathematical Statistics 43 (2), 651–655.

Esary, J. D., Proschan, F., Walkup, D. W., 1967. Association of Random Variables, with Applications. The Annals of Mathematical Statistics 38 (5), 1466–1474.

Fang, K.-T., Kotz, S., Ng, K.-W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London.

Farlie, D. J. G., 1960. The performance of some correlation coefficients for a general bivariate distribution. Biometrika 47, 307–323.

Faure, H., 1982. Discrépance de suites associées à un système de numération (en dimension s). Acta Arithmetica 41, 337Ű–351.

Feller, W., 1971. An Introduction to Probability Theory and its Applications. Vol. 2 of Wiley Series in Probability and Mathematical Statistics. Wiley, New York, NY.

Fermanian, J., Scaillet, O., 2003. Nonparametric estimation of copulas for time series. Journal of Risk 5, 25–54.

Fishburn, P., 1977. Mean-risk analysis with risk associated with below-target returns. Amercan Economic Review 67 (2), 116–126.

Forbes, K. J., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. Journal of Finance 57, 2223–2261.

Foulcher, S., Gourieroux, C., Tiomo, A., 2005. Latent Variable Approach to Modelling Dependence of Credit Risks: Application to French Firms and Implications for Regulatory Capital. Working Paper CREF 05-01, HEC, Montreal.

Frachot, A., Georges, P., Roncalli, T., 2001. Loss Distribution Approach for Operational Risk. Working Paper , Crédit Lyonnais.

Frachot, A., Georges, P., Roncalli, T., 2003. Loss Distribution Approach in Practice. Working Paper , Crédit Lyonnais.

Frachot, A., Roncalli, T., Salomon, E., 2004. The correlation problem in operational risk. Working paper, Crédit Lyonnais.

Frahm, G., Junker, M., Schmidt, R., 2005. Estimating the tail–dependence coefficient: Properties and pitfalls. Insurance: mathematics and Economics 37, 80–100.

Frank, M. J., Nelsen, R. B., Schweizer, B., 1987. Best–possible bounds for the distribution of a sum—a problem of Kolmogorov. Probability Theory and Related Fields 74 (2), 199–211.

Frank, M. J., Schweizer, B., 1979. On the duality of generalized infimal and supremal convolutions. Rendiconti di Matematica 12, 1–23.

Fréchet, M., 1951. Sur les tableaux de corrélation dont les marges sont donnés. Annales de l'Université de Lyon 3 (14), 53–77.

Frees, E. W., Valdez, E. A., 1999. Understanding relationships using copulas. North American Actuarial Journal 2 (1), 1–25.

Frey, R., McNeil, A., 2003. Dependence modelling, model risk and model calibration in models of portfolio credit risk. Journal of Risk 6 (1), 59–92.

Frey, R., McNeil, A. J., 2001. Modelling Dependent Defaults. Preprint, University and ETH Zürich.

Frey, R., McNeil, A. J., 2002. VaR and Expected Shortfall in Portfolios of Dependent Credit Risks: Conceptual and Practical Insights. Journal of Banking & Finance 26, 1317–1334.

Galambos, J., 1987. The Asymptotic Theory of Extreme Order Statistics. Krieger.

Gao, Y., Zhou, J., 2010. Tail dependence in international real estate securities markets. Journal of Real Estate Finance and Economics, 1–31.

Garcia, R., Tsafack, G., 2009. Dependence structure and extreme comovements in international equity and bond markets. Working Paper 2009s-21, Cirano.

Geffroy, J., 1958/59. Contribution à la théorie des valuers extrêmes. Publications de l'Institut de Statistique de l'Université de Paris 7/8, 37–184.

Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82 (3), 543–552.

Genest, C., MacKay, R., 1986. The joy of copulas: Bivariate distributions with uniform marginals. The American Statistician 40 (4), 280–283.

Genest, C., Nešlehová, J., 2007. A primer on copulas for count data. The ASTIN Bulletin 37, 475–515.

Genest, C., Quessy, J.-F., Rémillard, B., 2006. Goodness–of–fit procedures for copula models based on the integral probability transformation. Scandinavian Journal of Statistics 33, 337–366.

Genest, C., Rémillard, B., 2008. Validity of the parametric bootstrap for goodness–of–fit testing in semiparametric models. Annales de l'Institut Henri Poincaré—Probabilités at Statistiques 44 (6), 1096–1127.

Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness–of–fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics 44, 199–213.

Genest, C., Rivest, L. P., 1989. A characterization of Gumbel's family of extreme value distributions. Statistics and Probability Letters 8, 207–211.

Genest, C., Rivest, L. P., 1993. Statistical Inference Procedures for Bivariate Archimedean Copulas. Journal of the American Statistical Association 88, 1034–1043.

Gersbach, H., Lipponer, A., 2003. Firm defaults and the correlation effect. European Financial Management 9 (3), 361–377.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer.

Grubel, H. J., December 1968. Internationally diversified portfolios: Welfare gains and capital flows. American Economic Review 58, 1229–1314.

Gumbel, E. J., 1960. Distributions à plusieurs variables dont les marges sont données. Comptes Rendus de l'Académie des Sciences 246, 2717–2719.

Haas, M., Mittnik, S., Yener, T., 2009. Gering korrelierte Anlageklassen—Diversifikationsmodell der Vergangenheit? Studienreihe, Bayerisches Finanz Zentrum e.V., Munich.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2, 84–90, corrections, 1960, *ibid*, 2, 190.

Hansen, P. R., Lunde, A., 2005. A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? Journal of Applied Econometrics 20, 873–889.

Heckman, P. E., Meyers, G. G., 1983. The calculation of aggregate loss distributions from claim severity and claim count distributions. Proceedings of the Casualty Actuarial Society LXX, 22–61.

Henze, N., 1996. Empirical-distribution-function goodness-of-fit tests for discrete models. Canadian Journal of Statistics 24 (1), 81–93.

Hilliard, J., 1979. The relationship between equity indices on world exchanges. Journal of Finance 34 (1), 103–114.

Hlawka, E., 1961. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. Annali di Matematica Pura ed Applicata 54 (1), 325–333.

Höffding, W., 1940. Masstabinvariante Korrelationstheorie. Schriften des Mathematischen Instituts und des Instituts fur Angewandte Mathematik der Universität Berlin 5, 179–233.

Hougaard, P., 1986. A class of multivariate failure time distributions. Biometrika 73, 671–678.

Hu, L., 2006. Dependence patterns across financial markets: a mixed copula approach. Applied Financial Economics 16, 717–729.

Hult, H., Lindskog, F., 2002. Multivariate extremes, aggregation and dependence in elliptical distributions. Advances in Applied Probability 34 (3), 587–608.

Jin, X., Zhang, A., 2006. Reclaiming Quasi–Monte Carlo Efficiency in Portfolio Value–at–Risk Simulation Through Fourier Transform. Management Science 52 (6), 925–938.

Joag-Dev, K., Proschan, F., 1983. Negative association of random variables. Annals of Statistics 11, 286–295.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Vol. 73 of Monographs on Statistics and Applied Probability. Chapman & Hall, London, Weinheim, New York.

Joe, H., Xu, J., 1996. The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Jondeau, E., Poon, S.-H., Rockinger, M., 2006. Financial Modeling Under Non-Gaussian Distributions. Springer.

Juri, A., Wüthrich, M. V., 2002. Copula convergence theorems for tail events. Insurance: Mathematics and Economics 30, 405–420.

Kaplanis, E. C., 1988. Stability and forecasting of the comovement measures of international stock market returns. Journal of International Money and Finance 7, 63–75.

Karolyi, A., Stulz, R. M., 1996. Why do markets move together? An investigation of U.S.–Japan stock return comovements. Journal of Finance 51, 951–986.

Kimeldorf, G., Sampson, A. R., 1987. Positive dependence orderings. Annals of the Institute of Statistical Mathematics 39, 113–128.

King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62 (4), 901–933.

Koch, P. D., Koch, T. W., 1991. Evolution in dynamic linkages across daily national stock indexes. Journal of Internaitonal Money and Finance 10, 231–251.

Kocis, L., Whiten, W. J., 1997. Computational Investigations of Low–Discrepancy Sequences. ACM Transactions on Mathematical Software 23 (2), 266–294.

Koksma, J. F., 1942/43. Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1. Mathematica B (Zutphen) 11, 7–11.

Kolmogorov, A. N., 1933. Sulla determinazione empirica di una legge di distributione. Giornale dell'Instituto Italiano Attuari 4, 83–91.

Laha, R. G., Rohatgi, V. K., 1979. Probability Theory. John Wiley & Sons.

Lancaster, H. O., 1982. Dependence, measures and indices of. In: Encyclopedia of Statistical Sciences. Wiley, pp. 334–339.

Ledford, A. W., Tawn, J. A., 1996. Statistics for near independence in multivariate extreme values. Biometrika 83, 169–187.

Ledford, A. W., Tawn, J. A., 1997. Modeling dependence within joint tail regions. Journal of the Royal Statistical Society: Series B 59 (2), 475–499.

Lehmann, E. L., 1966. Some concepts of dependence. The Annals of Mathematical Statistics 37 (5), 1137–1153.

Levy, H., Sarnat, M., 1970. International diversification of investment portfolios. American Economic Review 60 (4), 668–675.

Lindskog, F., 2000. Linear correlation estimation. Research report, RiskLab Switzerland, `http://www.risklab.ch/ftp/papers/LinearCorrelationEstimation.pdf`.

Ling, C. H., 1965. Representations of associative functions. Publicationes Mathematicae Debrecen 12, 189–212.

Lintner, J., 1965. Security prices, risk and maximal gains from diversification. Journal of Finance 20, 587–615.

Longin, F., Solnik, B., 1995. Is the correlation in international equity returns constant: 1960–1990? Journal of International Money and Finance 14, 3–26.

Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance 56, 649–676.

Loretan, M., English, W. B., 2000. Evaluating "correlation breakdowns" during periods of market volatility. Working Paper 658, Board ofGovernors of the Federal Reserve System International Finance.

Makarov, G., 1981. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. Theory of Probability and its Applications 26, 803–806.

Maldonado, R., Saunders, A., 1981. International portfolio diversification and the inter–temporal stability of international stock market relationships, 1957–1978. Financial Management 10 (4), 54–63.

Malevergne, Y., Sornette, D., 2003. Testing the Gaussian Copula Hypothesis for Financial Assets Dependences. Quantitative Finance 3, 231–250.

Mardia, K. V., 1964. Asymptotic independence of bivariate extremes. Calcutta Statistical Association Bulletin 13, 172–178.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 56, 649–676.

Marshall, A. W., 1996. Copulas, marginals, and joint distributions. In: Rüschendorff, L., Schweizer, B., Taylor, M. D. (Eds.), Distributions with Fixed Marginals and Related Topics. Institute of Mathematical Statistics, Hayward, CA, pp. 213–222.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton Series in Finance. Princeton University Press.

Meijer, H. G., 1968. The Discrepancy of a g–adic Sequence. Indagationes Mathematicae 30, 54–66.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29, 449–470.

Mikosch, T., 2006. Copulas: Tales and Facts. Extremes 9 (1), 3–20.

Mittnik, S., Yener, T., 2009. Estimating Operational Risk Capital for Correlated, Rare Events. Journal of Operational Risk 4 (4), 1–23.

Morgenstern, D., 1956. Einfache Beispiele Zweidimensionaler Verteilungen. Mitteilungsblatt für Mathematische Statistik 8, 234–235.

Morokoff, W. J., Caflisch, R. E., 1994. Quasi-Random Sequences and their Discrepancies. SIAM Journal on Scientific Computing, 15 (6), 1251–1279.

Morokoff, W. J., Caflisch, R. E., 1995. Quasi–Monte Carlo Integration. Journal of Computational Physics 122, 218–230.

Mossin, J., 1966. Equilibrium in a capital market. Econometrica 34 (4), 768Ű–783.

Nelsen, R. B., 1999. An Introduction to Copulas. Vol. 139 of Lecture Notes in Statistics. Springer, New York.

Niederreiter, H., 1978. Quasi–Monte Carlo Methods and Pseudo–Random Numbers. Bulletin of the American Mathematical Society 84, 957–1041.

Niederreiter, H., 1992. Random Number Generation and Quasi–Monte Carlo Methods. Society for Industrial Mathematics.

Oakes, D., 1982. A model for association in bivariate survival data. Journal of the Royal Statistical Society. Series B 44, 414–422.

Owen, A. B., 1995. Randomly permuted $(t, m, s$–nets and (t, s)–sequences). In: Monte Carlo and Quasi–Monte Carlo Methods in Scientific Computing. Springer.

Owen, A. B., 2005. Multidimensional Variation for Quasi–Monte Carlo. In: Contemporary Multivariate Analysis and Design of Experiments. World Scientific, pp. 49–74.

Owen, A. B., 2006. Halton sequences avoid the origin. SIAM Review 48 (3), 487–503.

Panjer, H., 2006. Operational Risk: Modeling Analytics. Wiley.

Panjer, H. H., 1981. Recursive evaluation of a family of compound distributions. ASTIN Bulletin 12, 22–26.

Panton, D. B., Lessig, V. P., Joy, O. M., 1976. Comovement of international equity markets: A taxonomic approach. Journal of Financial and Qunatitative Analysis 11, 415–432.

Patton, A. J., 2004. On the out–of–sample importance of skewness and asymmetrics dependence for asset allocation. Journal of Financial Econometrics 2, 130–168.

Pearson, E. S., Stephens, M. A., 1962. The goodness-of-fit tests based on w_n^2 and u_n^2. Biometrika 49, 397–402.

Pfeifer, D., Nešlehová, J., 2004. Modeling and generating dependent risk processes for IRM and DFA. ASTIN Bulletin 34 (2), 333–360.

Pickands, J., 1981. Multivariate extreme value distributions. Bulletin of the International Statistical Institute 49, 859–878.

Rachev, S., Mittnik, S., 2000. Stable Paretian Models in Finance. Wiley.

Ratner, M., 1992. Portfolio diversification and the inter-temporal stability of international stock indices. Global Finance Journal 3 (1), 67–77.

Rényi, A., 1959. On measures of dependence. Acta Mathematica Academia Scientia Hungarica 10, 441–541.

Reshetar, G., 2008. Dependence of operational losses and the capital at risk. Technical Report, University of Zurich—Swiss Banking Institute (ISB).

Robertson, J., 1992. The computation of aggregate loss distributions. Proceedings of the Casualty Actuarial Society LXXIX, 57–133.

Roncalli, T., 2002. Gestion des risques multiples. Working paper, Crédit Lyonnais.

Ronn, E., Sayrak, A., Tompaidis, S., 2009. The impact of large changes in asset prices on intra–market correlations in the domestic and international markets. The Financial Review 44, 405–436.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. The Annals of Mathematical Statistics 23 (470–472).

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13 (3), 341–360.

Rubino, G., Tuffin, B., 2009. Rare Event Simulation Using Monte Carlo Methods. Wiley.

Salmon, F., 2009. Recipe for Disaster: The Formula That Killed Wall Street. Wired Magazine, 17.03.2009.

Sarkar, P. K., Prasad, M. A., 1987. A Comparative Study of Pseudo and Quasi Random Sequences for the Solution of Integral Equations. Journal of Computational Physics 68, 66–88.

Schucany, W. R., Parr, W. C., Boyer, J. E., 1978. Correlation structure in Farlie–Gumbel–Morgenstern distributions. Biometrika 65 (3), 650–653.

Schweizer, B., Sklar, A., 1961. Associative functions and statistial triangle inequalities. Publicationes Mathematicae Debrecen 8, 169–186.

Schweizer, B., Wolff, E., 1981. On nonparametric measures of dependence for random variables. Annals of Mathematical Statistics 11, 195–210.

Segers, J., 2004. Non–parametric inference for bivariate extreme–value copulas. Working Paper 2004Ű91, Tilburg University.

Shaked, M., Shanthikumar, J. G., 2006. Stochastic Orders. Springer.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19 (3), 425–442.

Sibuya, M., 1960. Bivariate extreme statistics. Annals of the Institute of Statistical Mathematics 11, 195–210.

Sklar, A., 1959. Fonctions de répartition a n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Universitá de Paris 8, 229–231.

Smirnov, N., 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bulletin Mathematique de l'Université de Moscou 2, 3–14.

Sobol, I. M., 1967. Distribution of points in a cube and approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7, 86–112.

Solnik, B., 1974. Why not diversify internationally rather than domestically? Financial Analysts Journal 30, 48–54.

Stute, W., Gonzáles-Manteiga, W., Presedo-Quindimil, M., 1993. Bootstrap based goodness–of–fit–tests. Metrika 40 (243–256).

Tchen, A. H., 1980. Inequalities for distributions with given marginals. Annals of Probability 8 (4), 814–827.

Tiago de Oliveira, J., 1962/63. Structure theory of bivariate extremes: extensions. Estudos de Matematica, Estatistica e Econometria 7, 165Ű195.

Tukey, J. W., 1958. A problem of Berkson, and minimum variance orderly estimators. Annals of Mathematical Statistics 29 (2), 588–592.

Watson, J., 1980. The stationarity of inter–country correlation coefficients: A note. Journal of Business Finance and Accounting 7.

Williamson, R. C., Downs, T., 1990. Probabilistic arithmetic: Numerical methods for calculating convolutions and dependency bounds. International Journal of Approximate Reasoning 4, 89–158.

Wozniakowski, H., 1991. Average case complexity of multivariate integration. Bulletin of the American Mathematical Society 24 (1), 185–193.