NTNU
Innovation and Creativity

Volatility and Dependence in Fixed Income Forward Rates with Application to Market Risk of Derivative Portfolios

Bjørn Erlend Vesterdal

Master of Science in Physics and Mathematics
Submission date: June 2006
Supervisor: Jacob Laading, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

Proper risk estimates are vital for decision support and for monitoring purposes when considering portfolios of financial assets. The purpose here is to investigate which models for the volatility and dependence in forward rates of the fixed income market are best suited to estimate the short term market risk of portfolios of interest rate derivatives.

Assignment given: 24. January 2006
Supervisor: Jacob Laading, MATH

Preface

The work on this thesis was carried out in the spring of 2006 over a period of 20 weeks, and the thesis itself counts 30 credits.

I was new to the field of quantitative risk management when I began the research process for this thesis, but it did not take long to make the pleasant discovery of the wide range of research that has been done on the application of different statistical tools to this field. Exploring these techniques has been both interesting and instructive, maybe in particular the copula approach to modeling dependence.

A range of tools have proved useful for the programming and analysis part of the work done here. The C++ program code was written in the `Bloodshed Dev-C++` IDE, which uses a Mingw compiler. For specialized statistical and numerical algorithms, the `Goose` and `Matpack` libraries have been used, as well as the *Numerical recipes* of Press et al. The `R` statistics environment was used for data analysis and graphics.

I would like to thank my supervisor Jacob Laading for the guidance and help he has provided throughout the work on the thesis. I would also like to thank Daniel Berg for giving me access to the `Copulae R` library and for helpful comments.

Trondheim, june 2006

Abstract

This thesis explores the modeling of volatility and dependence in forward rates in the fixed income market for the purpose of risk estimation in derivative portfolios. A brief background on popular quantile-based risk measures is given. A short introduction is given to GARCH-type volatility models, as well as copula and vine models for dependence between random variables. Some details on parameter estimation and sampling related to these models are also provided. A backtesting procedure is performed using various combinations of volatility and dependence models. The results of this procedure indicate that the Student's t copula is preferable among the dependence structures considered. Also, none of the choices of conditional distribution for the volatility models provide good results at all the percentiles considered, but the normal distribution appears to be a good choice far into the tails.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Fixed income derivative instruments make up a huge market in international financial services, and are present in substantial quantities in the portfolios of banks and other financial institutions. For a holder of such a portfolio, there can be considerable risk associated with movements in the interest rate, that is, the *market risk* associated with the instruments. A good model for this risk can be an invaluable tool for decision support and monitoring trading activities.

The value of a portfolio consisting of fixed income instruments may be dependent on the entire curve of short-, medium-, and long-term rates, known as the *yield curve*. Hence, so may the risk. In particular, instruments that provide some form of optionality are often valued using the *forward rates* associated with market quoted yields. The idea pursued here is that since these models require forward rates for pricing purposes, we will try to model the movements in forward rates for risk evaluation purposes as well. The approach taken is thus to try to recreate curves of future forward rates corresponding to typical actual movements in these rates, and use these curves and standard pricing approaches to obtain appropriate measures of risk.

We start off in chapter 2 with a look at Value at Risk (VaR), a very popular measure of portfolio risk. Some common approaches to the calculation of VaR are considered, and also an alternative risk measure suggested to correct some shortcomings of VaR, the Expected Shortfall (ES).

Chapter 3 considers possible ways to model the volatility of individual time series of financial asset data, through variations of $GARCH$-models. Chapter 4 takes a look at an increasingly popular approach to modeling the dependence between financial variables, through the use of *copulae*. Chapter 5 considers an extension of the copula approach known as a *vine*, allowing for a different specification of dependence. Finally, chapter 6 analyzes how the different model specifications behave when applied to our example of forward rates for fixed income derivative risk estimation.

Chapter 2

Value at Risk: Definitions and common assumptions

2.1 Definition

Value at Risk can be defined as how much a portfolio can decrease in value, within a given degree of confidence, over a specified time interval. To formalize this, first assume the following notation:

$$\begin{array}{rcl} S & = & \text{a vector of } n \text{ risk factors, typically market prices or rates} \\ \Delta t & = & \text{the risk measurement horizon} \\ V(S,t) & = & \text{portfolio value at time } t \text{ as a function of the risk factors } S \\ L & = & \text{portfolio profit or loss over period } \Delta t, \\ & = & -\Delta V = V(S,t) - V(S+\Delta S, t+\Delta t) \\ F_L(x) & = & P(L < x), \text{ the distribution of L.} \end{array}$$

The VaR for a portfolio is then a percentile of its loss distribution F_L.

Definition 2.1 *(Value at Risk) For a given confidence level $\alpha \in (0,1)$, the α-level VaR is the smallest number x such that the probability that the loss L exceeds x is no larger than $1-\alpha$. In the notation above,*

$$VaR = F_L^{-1}(\alpha) = \inf\{x : F_L(x) \geq \alpha\}.$$

Its estimator, which will depend on the assumptions we make, is then

$$\widehat{VaR} = \hat{F}_L^{-1}(\alpha) = \inf\{x : \hat{F}_L(x) \geq \alpha\}$$

for some estimate $\hat{F}_L$ of the loss distribution F_L.

For our example of a fixed income portfolio, the risk factors S would be a set of forward rates, representing the market risk from changes in interest rates. The timestep Δt in estimation of market risk is typically short, somewhere between one day and two weeks. As pointed out in Glasserman (2004) [35], two weeks is often interpreted as the time it might take to unwind complex positions in the case of an adverse market move.

2.2 Common assumptions for VaR calculations

2.2.1 The Delta-Normal method

A common approach to the calculation of VaR, introduced in RiskMetrics (1996) [59], is to assume that changes in the risk factors are normally distributed and that the portfolio value is linear in the risk factors. Under these assumptions, the possible changes in portfolio values, and thus the loss distribution of the portfolio, are given by a normal distribution as well.

To see how this works, begin with the assumption on the risk factors. Over a short time horizon, the mean of ΔS can be assumed to be negligible. Thus, we impose the distribution

$$\Delta S \sim N(0, \Sigma)$$

on the changes in risk factors for some covariance matrix Σ.

The linearity assumption means that

$$\Delta V = \delta^\top \Delta S \tag{2.1}$$

for some vector δ of portfolio value sensitivities to changes in risk factors. Thus the change in portfolio value is a linear combination of changes in risk factors, and the loss distribution is easily seen to be

$$L \sim N(0, \sigma_L^2)$$

with

$$\sigma_L^2 = \delta^\top \Sigma \delta.$$

The α-level VaR is then simply

$$VaR = \Phi^{-1}(\alpha)\sigma_L = \Phi^{-1}(\alpha)\sqrt{\delta^\top \Sigma \delta}$$

where Φ is the standard normal distribution.

This approach is fast, and also easy to implement given that portfolio deltas are readily available or easy to evaluate. It does, however, have two major shortcomings. One lies in the assumption of normally distributed changes in risk factors, the discussion of which will be postponed to a later section. The other, which will be addressed here, is the assumption of portfolio value changes as a linear function of changes in risk factors.

The linearity assumption (2.1) is appropriate for a portfolio of instruments whose value is linear or approximately linear in the underlying risk factors. The problem occurs when optionality is introduced to the portfolio. Figures 2.1 and 2.2 show the effect of the linearity assumption given normally distributed changes in rates. The delta-normal method will use a linear approximation for change in portfolio value as a function of rates, leading to a normal distribution for the possible portfolio values at the VaR horizon as well, as shown in figure 2.1.

If, however, we assume that the actual value of the portfolio is a nonlinear function of the rates, the situation might look like figure 2.2. Here we have

introduced a value function similar to that of a cap (a call option on the rate). This will result in a skewed distribution of portfolio values, something the delta-normal approximation will be unable to take into account. The size of this effect on the VaR estimate will depend on the amount of nonlinearity, or how "bent" the value function is, between the value today and the value corresponding to the VaR α-quantile.

The delta approach outlined above can be improved for nonlinear positions by including higher order terms, see e.g. the delta-gamma approximation of RiskMetrics (1996) [59].

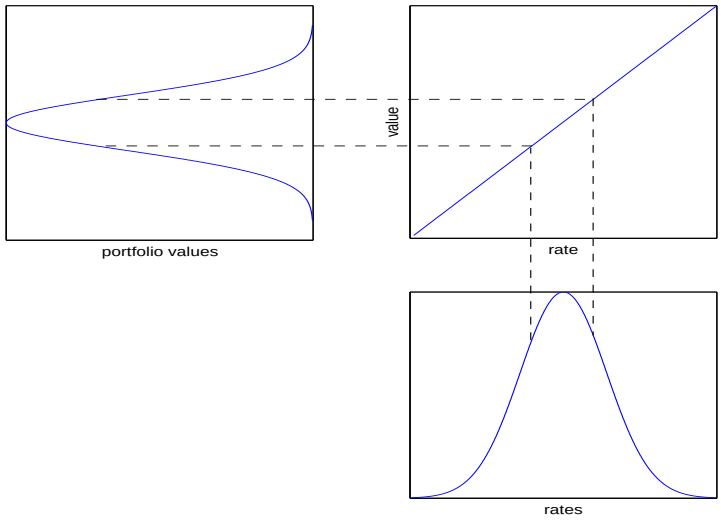


Figure 2.1: VaR estimation under the linearity assumption

2.2.2 Estimating VaR by simulation

The problems associated with the delta approximation can be circumvented by building the distribution of portfolio values through simulation. In its simplest form, a simulation proceeds as follows:

- For each of N replications
  - Generate a vector ΔS of changes in risk factors over time period Δt.
  - Revalue portfolio as $V(S + \Delta S, t + \Delta t)$.
- Sort the vector of portfolio values from previous step; select $V_{\lfloor N\alpha \rfloor}$ (the $N\alpha$-smallest value) as the α-level portfolio value
- Estimate VaR as $V(S,t) - V_{\lfloor N\alpha \rfloor}$

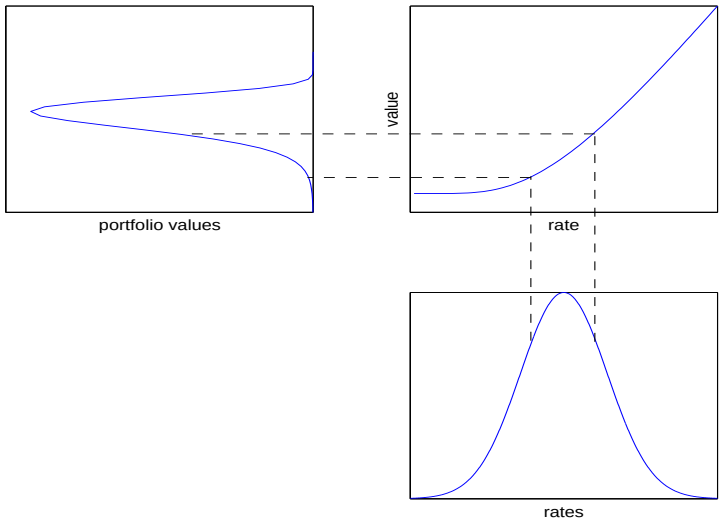


Figure 2.2: VaR estimation for an option portfolio

This is of course more time consuming than the delta-normal method, as we need N revaluations of the portfolio. Depending on the instruments contained in the portfolio, each revaluation may also be individually quite computationally expensive. On the plus side, nonlinearities in the portfolio value function are automatically taken into account. Also, we are free to choose the distribution from which to sample the vector ΔS of changes in risk factors, we are no longer limited to the normal distribution. As will become evident in later chapters, estimating distributional parameters and generating samples from the chosen distribution may also incur considerable computational effort. Note that the vector ΔS should be generated under the *real* measure, whereas the portfolio revaluations should be done under the *risk-neutral* measure.

A popular approach to VaR calculations is *historical simulation*. If a time window and a sampling frequency have been chosen, e.g. one year and daily sampling, this amounts to revaluing today's portfolio positions on the daily price changes over the past year, and choosing the appropriate percentile of the corresponding profit/loss-distribution as the VaR estimate. In terms of the above simulation algorithm, the changes in risk factors ΔS are taken as historical realizations over some time period. A few observations are made concerning common arguments for using historical simulation in Finger (2006) [29]. A sensible argument in favor of historical simulation is that the outcomes can be given a very intuitive interpretation; in historical simulation, the worst-case scenario corresponds to some historical event. It can be of interest in itself to know, for example, what the effect on the portfolio would be if an event corresponding to the worst event in the past year hit the market. It does not, however, necessarily make historical simulation a good forecasting technique. An argument that make less sense is that the model is assumption-free. Naturally, the model

implicitly assumes that the sampled historical data adequately represent the distribution of future returns and that the return distribution does not vary over the time period chosen. The historical simulation approach is easy to explain and implement, but those are perhaps its main virtues. The method will not be considered further here.

2.3 Properties of VaR as a risk measure

Before discussing the specific properties of VaR, we should have some notion of what constitutes a good risk measure. In Artzner et al. (1999) [7] four axioms are listed which should be fulfilled by a risk measure in order for it to be what they call *coherent*. Note that Artzner et al. (1999) [7] define risk through the future value of the portfolio, rather than the loss. To keep notation consistent with section 2.1, these axioms will be listed here as given in Frey & McNeil (2002) [32].

First, define a risk measure on a *convex cone* $\mathcal{M}$ of random variables. $\mathcal{M}$ being a convex cone means that if $L_1 \in \mathcal{M}$ and $L_2 \in \mathcal{M}$, then $L_1 + L_2 \in \mathcal{M}$, and also $\lambda L_1 \in \mathcal{M}$ for every $\lambda > 0$.

Definition 2.2 *(Risk measure) Given some convex cone $\mathcal{M}$ of random variables, a measure of risk with domain $\mathcal{M}$ is a mapping $\rho : \mathcal{M} \to \Re$.*

$\rho(L)$ can then be interpreted as the amount of capital that needs to be added to the portfolio with loss L to make the portfolio acceptable to some risk controller. The axioms of coherence are then the following.

Axiom 1 *(Translation invariance) For all $L \in \mathcal{M}$ and every $l \in \Re$ a translation invariant risk measure satisfies $\rho(L + l) = \rho(L) + l$.*

Axiom 2 *(Subadditivity) For all $L_1, L_2 \in \mathcal{M}$ a subadditive risk measure satisfies $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$.*

The subadditivity property implies that diversification does not increase risk. A risk measure which is not subadditive might, for instance, motivate breaking up a portfolio into smaller units to obtain a lower risk estimate for the sum of the smaller units than for the original portfolio. Economically, this does not make sense, we want the sum of the individual risks to provide an upper bound on the total portfolio risk.

Axiom 3 *(Positive homogeneity) For all $L \in \mathcal{M}$ and every $\lambda > 0$ a positive homogeneous risk measure satisfies $\rho(\lambda L) = \lambda \rho(L)$.*

Axiom 2 already implies $\rho(nX) \leq n\rho(X),\ n = 1, 2, ...$ Axiom 3 imposes the reverse inequality and requires equality for all positive λ. This means that combining identical positions with exposure to the same single risk factor into a larger portfolio does not give smaller risk for the resulting portfolio than the sum of the risks of the original positions.

Note that subadditivity and positive homogeneity together imply convexity, that is, the property $\rho(\lambda L_1 + (1 - \lambda)\rho(L_2)) \leq \lambda\rho(L_1) + (1 - \lambda)\rho(L_2), \lambda \in [0, 1]$.

Axiom 4 *(Monotonicity) For L_1, $L_2 \in \mathcal{M}$ such that $L_1 \leq L_2$ almost surely, a monotonic risk measure satisfies $\rho(L_1) \leq \rho(L_2)$.*

The definition of coherency of a risk measure follows.

Definition 2.3 *(Coherent risk measure) A risk measure ρ whose domain includes the convex cone $\mathcal{M}$ is called coherent (on $\mathcal{M}$) if it satisfies Axiom 1, 2, 3 and 4.*

VaR, as given in definition 2.1, can be seen to satisfy translation invariance, positive homogeneity and monotonicity. However, Artzner et al. (1999) [7] give examples to show that VaR does not, in, general, satisfy subadditivity. In other words, examples can be constructed where diversification increases risk as measured by VaR. Consider the following situation (Artzner et al. (1999) [7]):

- Two digital options A and B on a stock are available, both of which are dependent on the value S_T of the stock at time T.
- A pays 1000 if $S_T > U$ for a given U, nothing otherwise.
- B pays 1000 if $S_T < L$ for a given $L < U$, nothing otherwise.
- $\mathrm{P}\{S_T < L\} = \mathrm{P}\{S_T > U\} = 0.008$.

The 1% VaR of writing one of these options will be zero, whereas the 1% VaR of writing both options will be 1000, illustrating a situation where VaR is not subadditive (or, as the authors point out for this example, even convex).

Another criticism of VaR is that, since VaR represents an α-quantile, it can be interpreted as the *minimum* loss incurred with probability α. This is inconvenient for the assessment of risks that create large losses with a small probability. Jaschke (2002) [38] gives a few examples of strategies that can generate profits with a large probability, and large losses with a small probability:

- Increase the bet until a certain profit is reached (As in a doubling strategy).
- Buy defaultable bonds and sell riskless bonds.
- Sell far-out-of-the-money put options.
- Sell insurances/financial derivatives for rare events.

Another risk measure has been suggested to compensate for the problems of VaR, namely the *expected shortfall* (ES). In Frey & McNeil (2002) [32], ES is defined as

$$ES = \mathrm{E}(L|L > VaR). \tag{2.2}$$

ES is thus the expected loss, given that the loss exceeds VaR. Frey & McNeil (2002) [32] report that ES as defined in (2.2) is coherent for continuous loss distributions, which is sufficient for our purposes. See Acerbi & Tasche (2002) [6] for another definition of ES which is coherent under discontinuous loss distributions as well.

It can be noted that Embrechts, McNeil & Straumann (2002) [25] show that if the risk factors are *elliptically distributed* and the portfolios considered are *linear* in these factors, then VaR does in fact have the subadditivity property. Elliptical distributions are briefly discussed in section 4.1.1 below.

Chapter 3

Volatility specification

A realistic model for the volatility is crucial in obtaining sensible VaR estimates. Using historical data, we would want a model that reflects current conditions in the market and perhaps in some way captures typical movements in the volatility. However, as volatility is an unobservable and somewhat fickle quantity, the correct model choice is seldom obvious. There is a large number of ways to approach this problem, this chapter will review some of the simpler ones. The basic idea we pursue is to assume some model for the volatility of each separate risk factor, and not yet concern ourselves with the dependence between the factors.

3.1 The rectangular moving average

The perhaps simplest approach to volatility modeling is the rectangular moving average (RMA). Denoting by ϵ_t the return on the risk factor from time $t-1$ to t, the RMA volatility estimate is given by

$$\sigma_t^2 = \frac{1}{T}\sum_{i=0}^{T} \epsilon_{t-i}^2.$$

Under the assumption of a mean return of zero, this is just the sample variance. So, the RMA estimator is easy to evaluate and understand, but those are also the only advantages. Every return observation is weighted equally in this estimator, so the model will be slow to adjust to new market conditions. Sudden large positive or negative returns will have a limited effect on the volatility estimate, and then this effect will stay in the estimate until a time period T has passed. As noted in Franke, Härdle & Hafner (2004) [30], this typically leads to an underestimation when the market is moving towards a more volatile phase, and an overestimation when the market is moving towards a less volatile phase.

3.2 The exponentially weighted moving average

The exponentially weighted moving average (EWMA) estimator is an attempt to make the volatility estimate adjust more quickly to changing market conditions. For some starting value σ_{t-T}^2, this estimator is given by

$$\sigma_t^2 = (1-\lambda)\epsilon_{t-1}^2 + \lambda\sigma_{t-1}^2 = (1-\lambda)\sum_{i=0}^{T}\lambda^i\epsilon_{t-i}^2.$$

The name comes from the fact that the returns are given exponentially decreasing weights as you move backwards in time. This gives recent observations a larger weight, allowing the estimator to adjust to the current typical size of returns. The RMA estimator above is the boundary case of the EWMA estimator as $\lambda \to 1$, see Franke, Härdle & Hafner (2004) [30].

3.3 A GARCH model for the volatility

Financial data tend to show evidence of *volatility clustering*, that is, large (in absolute value) returns follow large returns, and small returns follow small returns. Thus, for a sufficiently long time series of financial data, one will typically see periods of large returns and periods of small returns. This indicates that the current volatility should depend in some way on previous volatility as well as previous returns. A class of models that attempts to capture this phenomenon, and which have become widespread in the industry, are the *GARCH*-models of Bollerslev (1986) [15].

The GARCH(p,q)-model is obtained by letting the innovations ϵ_t in the linear regression

$$y_t = x_t^\top b + \epsilon_t \tag{3.1}$$

follow the process

$$\begin{aligned} \mathrm{E}(\epsilon_t) &= 0, \\ \mathrm{V}ar(\epsilon_t) &= \sigma_t^2, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i\sigma_{t-i}^2, \end{aligned}$$

where

$$\begin{aligned} p &\geq 0,\ q > 0, \\ \alpha_0 &> 0, \\ \alpha_i &\geq 0,\ i = 1, ..., q, \\ \beta_i &\geq 0,\ i = 1, ..., p. \end{aligned}$$

In this model, we call the distribution of ϵ_t the *marginal* distribution, while the distribution of ϵ_t/σ_t is known as the *conditional* distribution.

Now, we need to choose values for p and q and also a sensible form of the regression in (3.1) to fit our risk factors in the fixed-income market. In a survey, Aas (2004b) [2] mentions the following simple GARCH(1,1)-model for the behaviour of the short-term interest rates:

$$
\begin{aligned}
r_t &= b_0 + (b_1 + 1)r_{t-1} + \epsilon_t, && (3.2)\\
\mathrm{E}(\epsilon_t) &= 0, && (3.3)\\
\mathrm{Var}(\epsilon_t) &= \sigma_t^2, && (3.4)\\
\sigma_t^2 &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2. && (3.5)
\end{aligned}
$$

Stationarity of the volatility in the model above is fulfilled for values of α_1 and β_0 such that $\alpha_1 + \beta_0 < 1$. Due to the first-order autoregression in (3.2), this model may be referred to as an AR(1)-GARCH(1,1) model.

The idea we pursue here is to try to model time series of individual forward rates applying over different maturities by a process of the form (3.2) - (3.5). This will be explored further in section 6.1.

For short-term interest rates, the volatility is known to exhibit a certain dependence on the interest rate level. Such models will not be applied here to our forward rates, see Aas (2004b) [2] for a review of several of these models. However, one other specification will be considered here, namely one which allows for asymmetry in the volatility.

An effect occasionally seen in financial data is that negative returns can tend to influence the level of volatility more than positive returns. A simple suggestion on how to allow the volatility model to account for this is given by Glosten, Jagannathan & Runkle (1993) [36]. The extension is to just add a term containing a dummy variable to (3.5), such that

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2 + \omega I(\epsilon_{t-1})\epsilon_{t-1}^2,$$

where

$$I(x) = \begin{cases} 1 & (x < 0) \\ 0 & (x > 0) \end{cases}$$

The parameter constraints are as in the GARCH(1,1) model, with the additional constraint

$$\alpha_1 + \omega > 0. \tag{3.6}$$

This way the volatility is allowed to react differently depending on the sign of the return. This model will be referred to as *GJR-GARCH*.

3.3.1 Choice of conditional distribution for the GARCH models

The standard choice of conditional distribution for a GARCH-model such as the one given in (3.2) - (3.5) is the normal distribution, that is,

$$f(\epsilon_t|\psi_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_t}\exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right),$$

where ψ_{t-1} denotes the information up to time $t-1$.

This yields a simple log-likelihood function to be maximized, namely

$$\log L = -\frac{1}{2}\sum_{t=1}^{T}\log\sigma_t^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{\epsilon_t^2}{\sigma_t^2},$$

excluding a constant term. This assumption on the conditional return distribution will give a marginal return distribution with heavier tails than the normal. This is a result of the returns being drawn with different volatilities, causing the return distribution to be a mixture of different normal distributions (see e.g. Duffie & Pan (1997) [24]). However, financial returns may have distributions which have even heavier tails than these.

An alternative is to use the standardized Student's t distribution as a conditional distribution, as suggested by Bollerslev (1987) [16]. He specifies this distribution as

$$f(\epsilon_t;\nu|\psi_{t-1}) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sigma_t\sqrt{\pi(\nu-2)}}\left[1+\frac{\epsilon_t^2}{(\nu-2)\sigma_t^2}\right]^{-(\nu+1)/2},$$

such that $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma_t^2$. The log-likelihood function is still relatively simple, but it now contains one more parameter, namely the degrees of freedom ν.

The Student's t distribution for the conditional returns handles excess kurtosis (that is, peakedness and heavy tails) in the data. It is, however, symmetric, and does not handle skewness. If we want the conditional distribution to handle skewness, we have an interesting problem. As explained in the introduction, the dependence structure of interest here will be modeled through copulae. As we will see below, such a model involves drawing dependent uniform random variables in the simulations. To preserve the dependence structure, achieving specific marginal distributions will generally have to be done through inversion of the distribution function. For the normal distribution and the Student's t distribution, efficient numerical schemes are available for such inversion, but such schemes may not be available for arbitrary skewed distributions. For example, we would not want to do a root-finding procedure on a numerical integration of the density function in every step in every dimension of the simulation procedure; while certainly possible, such an approach would be excrutiatingly slow. What we would like is a distribution which allows for skewness and excess kurtosis, while at the same time being relatively simple to invert.

A possibility is to use the inverse hyperbolic sine (IHS) distribution of Johnson (1949) [40]. The density is given by

$$f(\epsilon_t;\lambda,\delta|\psi_{t-1}) = \frac{s}{\sigma_t\delta\sqrt{2\pi\left[(\epsilon_t/\sigma_t s+\mu)^2+1\right]}}\exp\left(-\frac{[\sinh^{-1}(\epsilon_t/\sigma_t s+\mu)-\lambda]^2}{2\delta^2}\right),$$

where $-\infty<\lambda<\infty$, $\delta>0$, $s=\sqrt{\frac{1}{2}(\omega-1)(\omega\cosh(2\lambda)+1)}$, $\mu=\sqrt{\omega}\sinh(\lambda)$, and $\omega=\exp(\delta^2)$.

As the name implies, this density is the result of applying an inverse hyperbolic sine transformation to the standard normal distribution. It is applied with good reported results in Choi (2001) [20] as a conditional distribution for GARCH-models, and it is characterized there as "one of the most flexible distributions among all skewed and leptokurtic distributions". The IHS distribution can cover a wide range of values for skewness and kurtosis, and has the following simple inversion formula:

$$F^{-1}(u) = \frac{\sinh[\lambda + \delta\Phi^{-1}(u)] - \mu}{s},$$

where Φ is the standard normal distribution function. Thus, its inversion requires only the inversion of the standard normal distribution and a hyperbolic sine transformation.

A comprehensive test of volatility models with different conditional distributions is performed in Bao, Lee & Saltoğlu (2004) [8]. A wide variety of models and distributions are applied to density forecasting for two stock indices. For these data, the authors report that in terms of modeling the whole distribution, the normal and Student's t distribution perform well for one of the indices, while the normal distribution provides poor results in the tails for both indices. The IHS distribution does not rank among the best for any of their tests, but it is generally not among the worst either. Further, it is suggested that the choice of conditional distribution has a greater impact on results than the choice of volatility model for these two stock indices. A similar comparison of how our distributions and volatility models perform on the forward rate data will be given in chapter 6, and details on parameter estimation can be found in appendix A.2.

Chapter 4

Copulae and their applications to risk management

Following the approach in Embrechts, McNeil & Straumann (2002) [25], consider the real-valued random variables $(X_1, ..., X_n)$. The dependence between these variables is completely described by their joint distribution function

$$F(x_1, ..., x_n) = \mathrm{P}\{X_1 \leq x_1, ..., X_n \leq x_n\}.$$

The idea of copulae is to separate F into a part that describes the dependence structure and a part that describes the marginal distributions. Assume, for simplicity, that the marginal distributions $F_1, ..., F_n$ of F are continuous. Then the vector $\mathbf{X} = (X_1, ..., X_n)^\top$ can be transformed component-wise to have $U(0, 1)$ marginal distributions using the probability integral transform $T : \Re^n \to \Re^n, (x_1, ..., x_n)^\top \mapsto (F(x_1), ..., F(x_n))^\top$. The joint distribution function C of $(F_1(X_1), ..., F_n(X_n))^\top$ is then called the *copula* of $\mathbf{X}$ or of F.

More formally, we have the following theorem, known as *Sklar's theorem*, as stated in Embrechts, Lindskog & McNeil (2001) [26]:

Theorem 4.1 *(Sklar's theorem) Let F be an n-dimensional distribution function with margins $F_1, ..., F_n$. Then there exists an n-copula C such that for all $\mathbf{x} \in \Re^n$*

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)). \tag{4.1}$$

If $F_1, ..., F_n$ are continuous, then C is unique, otherwise C is uniquely determined on $\operatorname{Ran} F_1 \times \cdots \times \operatorname{Ran} F_n$. Conversely, if C is an n-copula and $F_1, ..., F_n$ are distribution functions, then the function F defined above is an n-dimensional distribution function with marginals $F_1, ..., F_n$.

$\operatorname{Ran} F_i$ here denotes the range of F_i. This theorem shows the appealing property that dependence and marginal distributions can be considered separately. For $F_1, ..., F_n$ continuous, the equation (4.1) can be stated the other way around:

Corollary 4.2 *Let F be an n-dimensional distribution function with continuous margins $F_1, ..., F_n$ and copula C, where C satisfies (4.1). Then, for any $\mathbf{u} \in [0,1]^n$*

$$C(u_1, ..., u_n) = F(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)). \tag{4.2}$$

With this in mind, consider the following definition:

Definition 4.3 *A copula is a multivariate distribution function of a random vector in $\Re^n$ with $U(0,1)$ marginals. Alternatively, a copula is any function $C : [0,1]^n \to [0,1]$ with the following three properties:*

1. *$C(u_1, ..., u_n)$ is increasing in each component x_i.*
2. *$C(1, ..., 1, u_i, 1, ..., 1) = u_i \ \forall \ i \in \{1, ..., n\}, \ u_i \in [0,1]$.*
3. *$\forall \ (a_1, ..., a_n), (b_1, ..., b_n) \in [0,1]^n$, where $a_i \leq b_i$ we have $\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+...+i_n} C(u_{1i_1}, ..., u_{ni_n}) \geq 0$, where $u_{j1} = a_j, \ u_{j2} = b_j \ \forall \ j \in 1, ..., n$.*

Stating that a copula is a multivariate distribution with $U(0,1)$ marginals is a somewhat "operational" definition, as it does not specify what is meant by a multivariate distribution function. For a discussion on this and a more detailed version of definition 4.3, see Embrechts, Lindskog & McNeil (2001) [26].

The density associated with the copula C is given by

$$c(u_1, ..., u_n) = \frac{\partial C(u_1, ..., u_n)}{\partial u_1 \cdot ... \cdot \partial u_n}, \tag{4.3}$$

and the density f of an n-dimensional distribution F can be obtained by

$$f(x_1, ..., x_n) = c(F_1(x_1), ..., F_n(x_n)) \prod_{i=1}^{n} f_i(x_i). \tag{4.4}$$

4.1 Some examples of copulae

First, a particularly simple, but conceptually important example is the copula corresponding to independent variables:

Example 4.4 (Product copula) The special case of independent random variables corresponds to the *product copula*, or *independence copula*

$$C_\Pi(u_1, ..., u_n) = u_1 \cdots u_n.$$

The rest of this section presents five other commonly used copulae. Two of them, the Gaussian and Student's t copulae, are *implicit* copulae, meaning their density $c(u_1, ..., u_n)$ is given by a known multivariate distribution. Further, they are both *elliptical*, a term explained in section 4.1.1.

The other three, the Clayton, Gumbel and Frank copulae, are *explicit* copulae, meaning that their distribution function $C(u_1, ..., u_n)$ has a closed-form expression. Also, all three of these copulae belong to a subclass known as *Archimedean* copulae.

4.1.1 Elliptical copulae

Elliptical copulae are the copulae associated with elliptical distributions, that is, distributions whose contours of equal density form ellipsoids. Before defining elliptical distributions, consider the following definition from Embrechts, McNeil & Straumann (2002) [25].

Definition 4.5 *(Spherical distribution) A random vector* $\mathbf{X} = (X_1, ..., X_n)^\top$ *has a spherical distribution if for every orthogonal map* $U \in \Re^{n\times n}$ *(i.e. U is such that* $UU^\top = U^\top U = I_{n\times n}$*)*

$$U\mathbf{X} \stackrel{d}{=} \mathbf{X}. \tag{4.5}$$

In other words, rotation of the coordinates leaves the distribution of $\mathbf{X}$ unchanged. We can also say that X is spherically distributed if and only if it admits the stochastic representation

$$\mathbf{X} \stackrel{d}{=} R\mathbf{U}, \tag{4.6}$$

where R is a nonnegative random variable and $\mathbf{U}$ is a random vector independent of R which is uniformly distributed on the unit hypersphere $S_{n-1} = \{\mathbf{s} \in \Re^n | \|\mathbf{s}\|_2 = 1\}$. Thus spherical distributions have densities that are constant on spheres, such that if X has density $f(\mathbf{x})$, then f is of the form $f(\mathbf{x}) = g(\mathbf{x}^\top\mathbf{x})$ for a non-negative function g. For the multivariate normal distribution we would have $R \sim \sqrt{\chi^2_n}$, whereas $R^2/n \sim F(n, \nu)$, where F denotes the Fisher distribution, would give the multivariate Student's t distribution. These spherical distributions would then be the distributions of random variables with zero linear correlation, which in the normal case corresponds to independence.

Turning to the elliptical distributions, these can be defined as the affine maps of the spherical distributions (Embrechts, McNeil & Straumann (2002) [25]):

Definition 4.6 *(Elliptical distribution) For an affine map* $T : \Re^n \to \Re^n$, $\mathbf{x} \mapsto A\mathbf{x} + \mu$, $A \in \Re^{n\times n}$, $\mu \in \Re^n$, *and given a spherically distributed variable* Y, $X = T(Y)$ *has an elliptical distribution.*

An elliptically distributed vector with parameters μ and Σ can be represented similarly as in (4.6) by

$$\mathbf{X} \stackrel{d}{=} \mu + RA\mathbf{U}, \tag{4.7}$$

where A is a matrix such that $AA^\top = \Sigma$ and R and $\mathbf{U}$ are as above. If Y has density $f(\mathbf{y})$ and $|A| \neq 0$, then the density g of X can be written

$$g(\mathbf{x}) = \frac{1}{\sqrt{|\Sigma|}} f((\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)). \tag{4.8}$$

Note that in one dimension the class of elliptical distributions coincides with the class of symmetric distributions. Elliptically distributed random vectors have certain convenient properties regarding linear combinations and marginal and conditional distributions, see Embrechts, McNeil & Straumann (2002) [25].

A range of elliptical distributions have been applied to financial problems. In addition to the Gaussian and Student's t distributions, Bradley & Taqqu (2001)

[18] list the the logistic distribution, scale mixture distributions and stable laws as examples.

The focus here will be on vectors of random variables whose dependence is described by the copula of an elliptical distribution, but with arbitrary marginals. These kinds of distributions have been termed *meta-elliptical* by Fang, Fang & Kotz (2002) [28]. Examples of such distributions and discussions on their properties are found in e.g. Fang, Fang & Kotz (2002) [28] and Abdous, Genest & Remillard (2005) [5].

Example 4.7 (Gaussian copula) The copula of the n-dimensional Gaussian, or normal, distribution with correlation matrix R is given by

$$C_R^{Ga}(\mathbf{u}) = \Phi_R^n(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n)), \tag{4.9}$$

where Φ_R^n denotes the standardized n-variate normal distribution function with correlation matrix R and Φ^{-1} is the inverse of the standard normal distribution. C_R^{Ga} is then called a Gaussian copula.

Since the multivariate standardized normal density is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|R|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top R^{-1}\mathbf{x}\right\},$$

(4.4) yields the density of the Gaussian copula as

$$c(\mathbf{u};\theta) = \frac{1}{|R|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top (R-I)\mathbf{x}\right\},$$

where I is the identity matrix and $\mathbf{x} = (\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n))^\top$.

Example 4.8 (Student's t copula) The copula of the n-dimensional Student's t-distribution is

$$C_{\nu,R}^t(\mathbf{u}) = t_{\nu,R}^n(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_n)), \tag{4.10}$$

where $t_{\nu,R}^n$ is the distribution of the n-variate standardized Student's t distribution with ν degrees of freedom and scale matrix R, and t_ν is the distribution of the standardized univariate Student's t distribution with ν degrees of freedom.

Again, we can find the density of the t-copula through the density of the multivariate standardized Student's t distribution

$$f(\mathbf{x}) = \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)(\nu\pi)^{n/2}|R|^{1/2}}(1 + \frac{1}{\nu}\mathbf{x}^\top R^{-1}\mathbf{x})^{-(\nu+n)/2},$$

and application of (4.4), yielding

$$c(\mathbf{u};\theta) = \frac{\Gamma((\nu+1)/2)[\Gamma(\nu/2)]^{n-1}(1+\mathbf{x}^\top R^{-1}\mathbf{x})^{-(\nu+n)/2}}{|R|^{1/2}[\Gamma((\nu+1)/2)]^n \prod_{i=1}^n (1+x_i^2/\nu)^{(\nu+1)/2}}, \tag{4.11}$$

where $\mathbf{x}_t = (t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_n))^\top$.

4.1.2 Archimedean copulae

Archimedean copulae can be defined as follows (Bouyé et al. (2000) [17]):

Definition 4.9 *(Archimedean copula) The copula given by*

$$C(u_1,...,u_n) = \begin{cases} \phi^{-1}(\phi(u_1)) + ... + \phi(u_n)), & \sum_{i=1}^{n} \phi(u_i) \leq \phi(0) \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

where $\phi(u)$ is a C^2 function such that $\phi(1) = 0$, $\phi'(u) < 0$ and $\phi''(u) > 0$ for all $0 \leq u \leq 1$, is called an Archimedean copula.

The function ϕ is called the *generator* of the copula. Further, the density of an Archimedean copula can be represented as (see Savu & Trede (2004) [60])

$$c(u_1,...,u_n) = \phi^{-1(n)}(\phi(u_1) + ... + \phi(u_n)) \prod_{i=1}^{n} \phi'(u_i), \tag{4.13}$$

where $\phi^{-1(n)}$ is the derivative of order n of the inverse of the generator.

Now, the specification (4.12) is quite restrictive. Any marginal distributions of $C(u_1,...,u_n)$ which have the same dimension will be identical, resulting in very limited flexibility in the dependence modeling. Embrechts, Lindskog & McNeil (2001) [26] discuss multivariate extensions of the bivariate Archimedean copula which allow for different generators to be used for different margins. As an example, in three dimensions the distribution C could be specified through

$$\begin{aligned} C(u_1,u_2,u_3) &= C_1(C_2(u_1,u_2),u_3) & (4.14) \\ &= \phi_1^{-1}(\phi_1 \circ \phi_2^{-1}(\phi_2(u_1) + \phi_2(u_2)) + \phi_1(u_3)), & (4.15) \end{aligned}$$

and this generalizes to higher dimensions, allowing for $n-1$ distinct generators for an n-variate distribution. For a detailed discussion of this, see Embrechts, Lindskog & McNeil (2001) [26].

Example 4.10 (Clayton copula) The generator of the Clayton copula and its inverse are given by

$$\phi(t) = (t^{-\delta} - 1),\ \phi^{-1}(x) = (x+1)^{-1/\delta}.$$

The distribution function of the n-dimensional Clayton copula can be represented as

$$C(u_1,...,u_n) = \left(\sum_{i=1}^{n} u_i^{-\delta} - d + 1 \right)^{-1/\delta},$$

$\delta \geq 0$, with corresponding density, according to (4.3) given by

$$c(u_1,...,u_n) = \delta^n \frac{\Gamma(\frac{1}{\delta} + n)}{\Gamma(\frac{1}{\delta})} \left(\prod_{i=1}^{n} u_i^{-\delta-1} \right) \left[\sum_{i=1}^{n} u_i^{-\delta} - n - 1 \right]^{-1/\delta-1}. \tag{4.16}$$

The Clayton copula is asymmetric, with a greater dependence for observations in the negative tail.

Example 4.11 (Gumbel copula) The Gumbel copula has

$$\phi(t) = (-\log(t))^{\delta},\ \phi^{-1}(x) = \exp(-x^{1/\delta}),$$

and the distribution can be given by

$$C(u_1, ..., u_n) = \exp\left[\left(-\sum_{i=1}^{n}(-\log u_i)^{\delta}\right)^{1/\delta}\right]. \tag{4.17}$$

The Gumbel copula is also asymmetric, but with the dependence being greater in the positive tail.

Example 4.12 (Frank copula) The generator and its inverse for the Frank copula are given by

$$\phi(t) = \log\left(\frac{\exp(-\delta t) - 1}{\exp(-\delta) - 1}\right),\ \phi^{-1}(x) = 1 + \frac{x - \log(1 - \exp(\delta) + \exp(\delta + x))}{\delta},$$

resulting in the distribution

$$C(u_1, ..., u_n) = -\frac{1}{\delta}\log\left(1 + \frac{\prod_{i=1}^{n}(\exp(-\delta u_i) - 1)}{(\exp(-\delta) - 1)^{n-1}}\right)$$

for $\delta \in \Re\backslash\{0\}$. The bivariate Frank copula is the only Archimedean copula which is radially symmetric, meaning that the Frank distribution of (U_1, U_2) is the same as of $(1 - U_1, 1 - U_2)$. See Genest (1987) [33] for more on this and other properties of the Frank copula.

For both the Gumbel and the Frank copula, the expression for the density becomes rather involved in higher dimensions. In both cases it can be derived from (4.13) and the formulas for the derivatives of the generators and their inverses listed in Savu & Trede (2004) [60].

4.2 Dependence concepts

4.2.1 Linear correlation

By far, the most commonly used notion of dependence between random variables is the *linear correlation*, also known as the *Pearson correlation*.

Definition 4.13 *(Linear correlation) The linear correlation between the random variables X and Y is given by*

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}},$$

where $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ *is the covariance of X and Y and* $\mathrm{Var}(\cdot)$ *denotes the variance.*

The linear correlation measures linear dependence, such that $Y = aX + b,\ a \in \Re\backslash\{0\}, b \in \Re \Leftrightarrow |\rho(X, Y)| = 1$, otherwise $|\rho(X, Y)| < 1$. The linear correlation is easy to calculate, and for the multivariate normal distribution it sums

up the entire dependence structure. This means that a linear correlation of zero implies independence of the normal variates. Unfortunately, this is the only case where this is true. Embrechts, McNeil & Straumann (2002) [25] give a simple example of the breakdown of linear correlation: consider $X \sim \mathcal{N}(0,1),\ Y = X^2$. The third moment of the normal distribution is zero, so the linear correlation is zero despite strong dependence. For an in-depth discussion of the properties of linear correlation, see Embrechts, McNeil & Straumann (2002) [25].

4.2.2 Other measures of dependence

Rank correlation

Several interesting dependence measures are based on the concept of concordance. Two observations $(x, y)^T$ and $(\tilde{x}, \tilde{y})$ from a vector $(X, Y)^T$ of continuous random variables are said to be *concordant* if $(x - \tilde{x})(y - \tilde{y}) > 0$, and *discordant* if $(x - \tilde{x})(y - \tilde{y}) < 0$. An example of a concordance based dependence measure is Kendall's tau:

Definition 4.14 *Kendall's tau for the random variables X and Y is given by*

$$\tau(X, Y) = \mathrm{P}\{(X - \tilde{X})(Y - \tilde{Y}) > 0\} - \mathrm{P}\{(X - \tilde{X})(Y - \tilde{Y}) < 0\},$$

where $(\tilde{X}, \tilde{Y})$ is an independent copy of (X, Y).

Kendall's tau is, in other words, the probability of concordance minus the probability of discordance. An interesting example for the interpretation of Kendall's tau is provided by Kruskal (1958) [44]:

> A rewording of the interpretation of τ is the following. Suppose that observations (X_1, Y_1) and (X_2, Y_2) are drawn but that only X_1 and X_2 are revealed to us at first by some kind of mechanical device. Suppose further that we agree to play a game wherein we predict $Y_1 < Y_2$ when $X_1 < X_2$ and $Y_1 > Y_2$ when $X_1 > X_2$. If our prediction turns out to be correct we win one dollar; if wrong, we lose one dollar. After prediction, the mechanical device reveals Y_1 and Y_2 and the payoff is made. Our expected gain in dollars is τ.

In terms of the copula of X and Y, Kendall's tau can be expressed as

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) \mathrm{d}C(u, v) - 1.$$

A useful result is that, for elliptical copulae,

$$\tau(X, Y) = \frac{2}{\pi} \arcsin(\rho(X, Y)). \tag{4.18}$$

This provides a convenient way to estimate correlation parameters for elliptical copulae which may or may not have elliptical margins.

The following procedure for estimating Kendall's tau for a set of multivariate observations is given in Lindskog (2000) [48]. Consider T bivariate observations $\{(x_t, y_t)\}_{t=1}^T$. Kendall's tau is the probability of concordance minus the probability of discordance, so the sample version is

$$\hat{\tau} = \frac{c - d}{c + d}, \tag{4.19}$$

where c is the number of concordant pairs in the sample and d is the number of discordant pairs. In other words, for every (T+1)T/2 pairs of observations $(x_i, y_i), (x_j, y_j)$, increment c if $(x_i - x_j)(y_i - y_j) > 0$, and increment d if $(x_i - x_j)(y_i - y_j) < 0$.

A slight correction to (4.19) may be necessary in case of ties in the data, that is $x_i = x_j$ or $y_i = y_j$. In his explanation of the procedure, Lindskog (2000) [48] calls a case of $x_i = x_j$ an "extra y pair", denoted e_y, and conversely for e_x. The corrected estimator then becomes

$$\hat{\tau} = \frac{c - d}{\sqrt{c + d + e_y}\sqrt{c + d + e_x}}. \tag{4.20}$$

For multivariate observations with dimension $n > 2$, $\hat{\tau}$ would be calculated for every bivariate pair.

Another concordance-based dependence measure is Spearman's rho:

Definition 4.15 *Let (X, Y), $(\tilde{X}, \tilde{Y})$ and (X', Y') be independent copies. Spearman's rho is then given by*

$$\rho_S(X, Y) = 3(\mathrm{P}\{(X - \tilde{X})(Y - Y') > 0\} - \mathrm{P}\{(X - \tilde{X})(Y - Y') < 0\}). \tag{4.21}$$

Spearman's rho can also be expressed in terms of the copula C of (X, Y), by

$$\rho_S(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) \mathrm{d}u\mathrm{d}v - 3. \tag{4.22}$$

The relation (4.22) can be used to show that if X and Y are random variables with distribution functions F_1 and F_2, then

$$\rho_s(X, Y) = \rho(F_1(X), F_2(Y)), \tag{4.23}$$

where ρ denotes linear correlation, see Embrechts, Lindskog & McNeil (2001)[26]. Similarly to the relation (4.18) for Kendall's tau, we have that for the normal distribution Spearman's rho and the linear correlation are related through

$$\rho_S(X, Y) = \frac{6}{\pi} \arcsin\left(\frac{\rho(X, Y)}{2}\right). \tag{4.24}$$

where again ρ denotes linear correlation. Contrary to (4.18) however, the relation (4.24) does not hold in general for elliptical distributions, see Hult & Lindskog (2001) [37].

Tail dependence

Tail dependence is a measure of joint extreme co-movements of bivariate random variables, and is defined as follows.

Definition 4.16 *Let (X, Y) be continuous random variables with distribution functions F_1 and F_2. The coefficient of upper tail dependence of (X, Y) is*

$$\lim_{u \nearrow 1} P\{Y > F_2^{-1}(u) | X > F_1^{-1}(u)\} = \lambda_U,$$

provided $\lambda_U \in [0, 1]$ exists. If $\lambda_U \in (0, 1]$, X and Y are said to be asymptotically dependent in the upper tail, whereas if $\lambda_U = 0$, X and Y are said to be asymptotically independent in the upper tail.

Lower tail dependence is defined similarly.

Tail dependence can be shown to be a copula property, see Embrechts, McNeil & Straumann (2002)[25]. The Gaussian copula has asymptotic independence in both tails, while the Student's t copula has asymptotic dependence in both tails, except for the limiting case $\nu = \infty$, which gives the Gaussian copula (Embrechts, Lindskog & McNeil (2001) [26]). For general results on tail dependence in elliptical distributions, see Hult & Lindskog (2001) [37].

For Archimedean copulae, upper tail dependence has a simple expression in terms of the generator, given that the inverse of the generator is the Laplace transform of a strictly positive random variable. This expression is

$$\lambda_U = 2 - 2 \lim_{s \searrow 0} \left(\frac{\phi^{-1\prime}(2s)}{\phi^{-1\prime}(s)} \right),$$

and lower tail dependence is defined similarly, see Embrechts, Lindskog & McNeil (2001) [26]. The above mentioned additional condition on ϕ is fulfilled by the copulae considered here. The Clayton copula has lower tail dependence but not upper tail dependence, the Gumbel copula has upper tail dependence but not lower tail dependence, while the Frank copula has neither upper nor lower tail dependence, see Embrechts, Lindskog & McNeil (2001) [26].

4.3 Parameter estimation for selected copulae

The parameter estimation procedures that will be considered here are variations of maximum likelihood estimation (MLE), as outlined in Bouyé et al. (2000) [17]. Denote by Θ the parameter space, and let $\theta \in \Theta$ be a vector of parameters to be estimated. Let $L_t(\theta)$ be the likelihood for observation t. The exact maximum likelihood (EML) estimator of θ is then given by

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \sum_{t=1}^{T} \log L_t(\theta).$$

For estimating the parameters of an n-dimensional distribution F, the expression (4.4) allows us to write this expression as

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \sum_{t=1}^{T} \log c(F_1(x_1^t), ..., F_n(x_n^t)) + \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i(x_i^t), \tag{4.25}$$

or, assuming uniform margins,

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \sum_{t=1}^{T} \log c(u_1, ..., u_n). \tag{4.26}$$

In higher dimensions, the maximization procedure above can become quite involved. An alternative approach, treated in Joe (1997) [39], suggests estimating marginal distribution parameters and copula parameters separately. In the notation of (4.25) this would amount to estimating the parameters of each f_i

first, then using these in (4.25) to estimate the copula parameters. This method, called *inference function for margins* (IFM), does not in general give the same estimates as EML, but is reported to be much more efficient (Joe (1997) [39]). Further, it relies on sensible distributions being chosen for the margins to obtain reasonable copula parameters. To circumvent this difficulty, one can instead find the empirical copula marginals $\hat{u}_1, ..., \hat{u}_n$ and use these in (4.26) to estimate the copula parameters directly. Such a method is called *canonical maximum likelihood* (CML), and is outlined below. It is illustrated for the Gaussian copula in Bouyé et al. (2000) [17] and for the t-copula in Mashal & Zeevi (2002) [51], as shown below.

4.3.1 The empirical marginal transformation

Let $\mathcal{X}_T = \{\mathbf{X}_t\}_{t=1}^T$, $\mathbf{X}_t = (X_{t1}, ..., X_{tn})$ be our sample, assumed to be mutually independent and distributed according to some function F with marginals $F_1, ..., F_n$. If the marginals were known, we would have that $(U_1, ..., U_n) = (F_1(X_{t1}), ..., F_n(X_{tn})) \sim C(\cdot|\theta)$, with C being the copula of F.

However, the marginals are not known. Instead we construct the empirical marginals

$$\hat{F}_i(\cdot) = \frac{1}{T}\sum_{t=1}^{T} I(X_{ti} \leq \cdot),\ i = 1, ..., n \tag{4.27}$$

where $I(\cdot)$ is an indicator function. Then define

$$\hat{\mathbf{U}}_t = (\hat{F}_1(X_{t1}), ..., \hat{F}_n(X_{tn})), \tag{4.28}$$

and let $\mathcal{U}_T = \{\hat{\mathbf{U}}_t\}_{t=1}^T$ denote the *pseudo-sample*. In practice, the empirical marginals $\hat{F}_i(\cdot)$ are scaled by a factor of $\frac{T}{T+1}$ to keep the elements of the pseudo-sample from becoming 1.

4.3.2 Parameter estimation examples

Example 4.17 (Parameter estimation for the Gaussian copula) A simple method for parameter estimation for the Gaussian copula is shown in Aas (2004a) [1]. The only parameter to be estimated here is the correlation matrix R. However, to avoid a maximization procedure which can be computationally very expensive in high dimensions, Aas (2004a) [1] suggests an approximate solution through instead maximizing over the set of possible covariance matrices Σ. This problem has an analytical solution given by

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^\top,$$

where again $\mathbf{x}_t = (\Phi^{-1}(u_{t1}), ..., \Phi^{-1}(u_{tn}))^\top$. The u_t may be generated through the empirical marginal transformation in (4.27) and (4.28), in which case $\hat{\Sigma}$ would be a CML estimate, or from assumed specified marginals, which would lead to an IFM estimate. The correlation matrix R is then obtained through

$$\hat{R} = \Delta^{-1}\hat{\Sigma}\Delta^{-1},$$

Δ being a diagonal matrix of the square roots of the diagonal elements of $\hat{\Sigma}$.

Example 4.18 (Parameter estimation for the Student's t copula) A CML algorithm for parameter estimation for the t-copula is suggested in Mashal & Zeevi (2002) [51]. First, assume a pseudo-sample $\mathcal{U}_T$ has been generated by (4.27) and (4.28). The simultaneous maximization over both R and ν would be quite involved. As an alternative, Mashal & Zeevi (2002) [51] suggest exploiting the relation (4.18) to estimate R and then maximizing the likelihood only over ν. Estimating Kendall's tau from the procedure mentioned in section 4.2.2 and applying (4.18) for every bivariate pair yields an estimate $\hat{R}$ of the correlation matrix R. The maximization step now simplifies to

$$\hat{\nu} = \arg \max_{\nu \in (2,\infty]} \left[\sum_{t=1}^{T} \log c(\mathbf{u}_t; \nu, \hat{R}) \right], \tag{4.29}$$

with $c(\cdot)$ as in (4.11) and $\hat{R}$ fixed, a far more pleasant task than simultaneous maximization over both R and ν.

Details on the estimation of ν, as well as on parameter estimation for the Clayton and Frank copulae, can be found in appendix A.3.

4.4 Goodness-of-fit tests for copulae

When choosing a specific copula to model the dependence structure for a set of observations, it would be nice to get some idea how how well it does the job. A quite recent approach to this problem is through goodness-of-fit (GOF) tests. A comparative study of several such tests that have been proposed can be found in Berg & Bakken (2006) [11]. Several of the tests they consider are based on applying the probability integral transformation of Rosenblatt (1952) [57], then projecting the multivariate problem to a univariate problem, and finally using some univariate test statistic.

4.4.1 A test procedure based on the probability integral transformation

The probability integral transformation (PIT) is a method which converts any set of data with known distribution into a set of iid $U(0,1)$ variables. This section will describe how to apply this to copula GOF tests, as explained in Breymann, Dias & Embrechts (2003) [19].

Let $\mathbf{X} = (X_1, ..., X_n)$ be a random vector with continuous cdf $F(x_1, ..., x_n)$, and let $F_{X_i}(x_i) = P(X_i \leq x_i)$ $i = 1, ..., n$ denote the marginal distributions. Consider the n transformations

$$\begin{aligned}
T(x_1) &= P(X_1 \leq x_1) = F_{X_1}(x_1), \\
T(x_2) &= P(X_2 \leq x_2 | X_1 = x_1) = F_{X_2|X_1}(x_2|x_1), \\
&\vdots \\
T(x_n) &= P(X_n \leq x_n | X_1 = x_1, ..., X_{n-1} = x_{n-1}), \\
&= F_{X_n|X_1,...,X_{n-1}}(x_1|x_1, ..., x_{n-1}).
\end{aligned}$$

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

If C is the copula of $\mathbf{X}$, such that

$$F(x_1, ..., x_n) = C(F_{X_1}(x_1), ..., F_{X_n}(x_n)),$$

then let $C_i(u_1, ..., u_i)$ denote the joint i-marginal distribution:

$$C_i(u_1, ..., u_i) = C(u_1, ..., u_i, 1, ..., 1),\ i = 2, ..., n$$

of $U_1, ..., U_i$, with $C_1(u_1) = u_1$ and $C_n(u_1, ..., u_n) = C(u_1, ..., u_n)$.

The conditional distribution of U_i given $U_1, ..., U_{i-1}$ is then

$$C_i(u_i|u_1, ..., u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, ..., u_i)}{\partial u_1 \cdots \partial u_{i-1}} \Big/ \frac{\partial^{i-1} C_{i-1}(u_1, ..., u_{i-1})}{\partial u_1 \cdots \partial u_{i-1}},\ i = 2, ..., n,$$

which means we can write

$$Z_i = C_i(F_{X_i}(X_i)|F_{X_1}(X_1), ..., F_{X_{i-1}}(X_{i-1})),\ i = 2, ..., n.$$

This means that if our random variables $(F_{X_1}(x_1), ..., F_{X_n}(x_n))$ have the distribution function C, then $\Phi^{-1}(Z_i),\ i = 1, ..., n$ are iid $\mathcal{N}(0, 1)$, and our test statistic $S = \sum_{i=1}^{n} [\Phi^{-1}(Z_i)]^2$ has a chi-square distribution with n degrees of freedom.

Example 4.19 (Gaussian PIT) Suppose we have a set of data $\mathcal{X}_T = \{\mathbf{X}_t\}_{t=1}^T$ for which we want to test the hypothesis $\mathcal{H}_0$ that the dependence structure is Gaussian. Since we specifically want to test the dependence structure, the marginal distributions $F_{X_1}, ..., F_{X_n}$ of (4.30) are taken to be the empirical marginals.

The following procedure is then applied to each $\mathbf{X}_t \in \mathcal{X}_T$, assuming a mean of zero:

- Set $\mathbf{U}_t = (\hat{F}_1(X_{t1}), ..., \hat{F}_n(X_{tn})),\ t = 1, ..., T$, with $\hat{F}_i$ as in (4.27). If $\mathcal{H}_0$ is correct, then $\mathbf{U}_t \sim C_{Ga}$.
- Set $\mathbf{Y}_t = (\Phi^{-1}(U_{t1}), ..., \Phi^{-1}(U_{tn})),\ t = 1, ..., T$.
- Find the empirical covariance matrix Σ of $\mathcal{Y}_t = \{\mathbf{Y}_t\}_{t=1}^T$, and a matrix C such that $C^\top C = \Sigma$.
- Set $\mathbf{W}_t = C^{-1}\mathbf{Y}_t,\ t = 1, ..., T$. If $\mathcal{H}_0$ is correct, then the components of each $\mathbf{W}_t$ should be standard normally distributed and independent.
- Set $\mathbf{Z}_t = (\Phi^(W_{t1}), ..., \Phi^(W_{tn})),\ t = 1, ..., T$. Under $\mathcal{H}_0$, the components of each $\mathbf{Z}_t$ should be uniformly distributed on $[0, 1]^n$ and independent.

Comparison of the above procedure with the simulation procedure in section 4.7 should highlight why the PIT has been called the inverse of simulation; rather than drawing independent random variables and trying to transform them to variates from a given multivariate distribution, a multivariate distribution is assumed on the data and the procedure attempts to transform them to independent random variables.

Based on the recommendations for choosing a test procedure given in Berg & Bakken (2006) [11], we consider the test by Berg & Bakken (2005) [12]. Continuing to denote by Z_i, $i = 1, ..., n$ the variables resulting from applying the PIT to a multivariate vector $\mathbf{X} \in \Re^n$, this test uses the following transformation:

$$Z_i^* = \mathrm{P}(r_i \leq \tilde{Z}_i | r_1, ..., r_{i-1}) = \left(1 - \left(\frac{1 - \tilde{Z}_i}{1 - r_{i-1}} \right)^{n-(i-1)} \right), \tag{4.30}$$

where $\tilde{\mathbf{Z}}$ is the sorted version of $\mathbf{Z}$ and the r_i are the rank variables of $\mathbf{Z}$. The next step is to let

$$Y = \sum_{i=1}^{n} \gamma(Z_i; \alpha) \cdot \Phi^{-1}(Z_i^*)^2, \tag{4.31}$$

where γ is a weight function that adds flexibility to the test; for instance, choosing $\gamma(Z_i; \alpha) = (Z_i - 1/2)^\alpha$, $\alpha \in (2, 4, ...)$ results in a test which emphasizes the tails of the copula. (Berg & Bakken (2005) [12] refer to this choice of γ as *power tail weighting*). If the cdf of Y is denoted $F_Y(\cdot)$, the test is defined by the cdf $B(w)$ of $F_Y(Y)$, that is

$$B(w) = \mathrm{P}\{F_Y(Y) \leq w\}, w \in [0, 1]. \tag{4.32}$$

Under $\mathcal{H}_0$, $B(w) = w$.

In practice, for T observations of the vector Z, its empirical counterpart

$$\hat{B}(w) = \frac{1}{T+1} \sum_{t=1}^{T} I(F_Y(Y) \leq w), \; w = \frac{1}{T+1}, ..., \frac{T}{T+1}. \tag{4.33}$$

is used, where I is an indicator function. Some test statistic is then applied to $\hat{B}$. Here, the Anderson-Darling statistic is used, defined as

$$\mathcal{T}^{AD} = T \int_0^1 \frac{(\hat{F}(z) - F(z))^2}{F(z)(1 - F(z))} \mathrm{d}F(z).$$

For $F(z) = z$, an empirical version of this statistic is (see Berg & Bakken (2006) [11])

$$\hat{\mathcal{T}}^{AD} = \frac{T}{T+1} \sum_{t=1}^{T} \frac{(\hat{F}(\frac{t}{T+1}) - \frac{t}{T+1})^2}{\frac{t}{T+1}(1 - \frac{t}{T+1})}.$$

This procedure is then repeated for a suitable number of iterations using simulated values from the copula assumed under $\mathcal{H}_0$ in place of the observed data. Finally, the p-value is found by applying the empirical distribution of the test statistics of the simulated values to the test statistic from the observed data.

4.5 Simulation from copulae

4.5.1 Elliptical copulae

Example 4.20 (Simulation from the Gaussian copula) Simulating random variates from the Gaussian copula is simple. If $\mathbf{X}$ has a standardized multivariate normal distribution (that is, it has mean 0 and covariance equal to the correlation matrix R), then $\mathbf{X}$ can be represented as

$$\mathbf{X} \stackrel{d}{=} A\mathbf{Z},$$

where A is a matrix such that $AA^\top = R$ and $Z_1, ..., Z_n \sim \mathcal{N}(0,1)$ and independent. Drawing a vector $\mathbf{u}$ from the copula C_R^{Ga} in (4.9) can then be done by the following algorithm.

- Find a matrix A such that $AA^\top = R$.
- Draw a vector z of independent normal random variates from $\mathcal{N}(0,1)$.
- Set $\mathbf{y} = A\mathbf{z}$.
- Set $u_i = \Phi(y_i),\ i = 1, ..., n$,

yielding a vector $(u_1, ..., u_n) \sim C_R^{Ga}$.

Example 4.21 (Simulation from the Student's t copula) Simulating variates from the t-copula $C_{\nu,R}^t$ in (4.10) is also straightforward. If $\mathbf{X}$ can be represented as

$$\mathbf{X} \stackrel{d}{=} \frac{\sqrt{\nu}}{\sqrt{S}} A\mathbf{Z},$$

where $S \sim \chi_\nu^2$, A is a matrix such that $AA^\top = R$ and $Z_1, ..., Z_n \sim \mathcal{N}(0,1)$ are independent, then $\mathbf{X}$ has an n-dimensional standardized t_ν-distribution with ν degrees of freedom and scale matrix R.

An n-dimensional vector of variates from $C_{\nu,R}^t$ can then be drawn using the following algorithm.

- Find a matrix A such that $AA^\top = R$.
- Draw a vector $\mathbf{z}$ of independent normal random variates from $\mathcal{N}(0,1)$.
- Draw a random variate s from χ_ν^2 independent of $z_1, ..., z_n$.
- Set $\mathbf{y} = A\mathbf{z}$.
- Set $\mathbf{x} = \frac{\sqrt{\nu}}{\sqrt{s}}\mathbf{y}$.
- Set $u_i = t_\nu(x_i),\ i = 1, ..., n$.

The resulting vector is then $(u_1, ..., u_n) \sim C_{\nu,R}^t$.

4.5.2 Archimedean copulae

A general methodology that can be used for sampling many types of Archimedean copulae is suggested by Marshall & Olkin (1988) [50], and is described here as recounted in Aas (2004a) [1].

Let ϕ denote the generator of an Archimedean copula as in (4.12). If ϕ is the inverse of the Laplace transform of a distribution function G on $\Re^+$ such that $G(0) = 0$, then simulation from the copula can be done by means of the following algorithm:

- Draw a random variate $X \sim G$.
- Draw n independent uniform random variates $V_1, ..., V_n$.
- Set $\mathbf{U} = (\phi^{-1}(-\log(V_1)/X), ..., \phi^{-1}(-\log(V_n)/X))^\top$.

The $(n \times 1)$ vector $\mathbf{U}$ then has the distribution of the copula in question.

Example 4.22 (Simulation from the Clayton copula) For the Clayton copula, the inverse of the generator is the Laplace transform of a gamma variate X, such that $X \sim Ga(1/\delta, 1)$. The gamma distribution with arbitrary real positive shape parameter can be a little intricate to sample; the implementation here uses the `GammaDistribution` routine from the `matpack` [1] library. See Devroye (1986) [23] for a list of generators for the gamma distribution.

Example 4.23 (Simulation from the Gumbel copula) For the Gumbel copula, we would have $X \sim St(1/\delta, 1, \gamma, 0)$, where $St(\cdot)$ denotes the positive stable distribution and $\gamma = (\cos(\frac{\pi}{2\delta})^\delta)$. Aas (2004a) [1] recounts an algorithm, attributed to Nolan, for sampling positive stable variates.

Example 4.24 (Simulation from the Frank copula) Finally, for the Frank copula we would draw X from the logarithmic series distribution on the positive integers, with parameter $\alpha = (1 - \exp(-\theta))$. This discrete distribution has pmf

$$\mathrm{P}(X = i) = -\frac{\alpha^i}{i \log(1 - \alpha)}, \tag{4.34}$$

and Devroye (1986) [23] gives the following one-line generator, attributed to Kendall and to Kemp: Let $U, V \sim U[0, 1]$ Then

$$X \leftarrow \left\lfloor 1 + \frac{\log(V)}{\log(1 - (1 - \alpha)^U)} \right\rfloor \tag{4.35}$$

has the logarithmic series distribution with parameter α. Devroye (1986) [23] lists other generators reported to be faster, but the gain would be negligible for our implementation, so we will stick with (4.35).

Discussions of the above sampling procedures and some further details can be found in Frees & Valdez (1998) [31] and Melchiori (2006) [52].

[1] Matpack C++ Numerics and Graphics Library: www.matpack.de

Chapter 5

Vines

Vines are a type of graphical models introduced by Bedford & Cooke (2002) [10] to model dependence in high-dimensional distributions. A vine is a form of generalization of a Markov tree to allow for conditional dependence between pairs of variables. A detailed discussion on Markov trees and belief networks will not be given here, see e.g. Ripley (1996) [58]. However, a short introduction is in order.

First, some terminology:

- A *graph* is a collection of *vertices* and *edges*, where the edges represent pairs of vertices.
- A *path* on a graph is a list of vertices where each successive pair of vertices are joined by an edge.
- If every pair of vertices are joined by a path, the graph is said to be *connected*.
- A *cycle* is a path which returns to its origin without visiting any one vertice more than once.
- A *tree* is a connected graph with no cycles
- A graph is *directed* if the edges are ordered pairs of vertices commonly represented by arrows, otherwise it is *undirected*. In a directed graph, we might call the first vertex a *parent*, and the one(s) following it a *child*.

A commonly encountered type of graph is the *directed acyclic graph*, or *DAG*. Such a graph is often used for modeling causality, and if the vertices are random variables, it is often called (among other things) a *belief network*. A belief network expresses a natural factorization of the joint distribution of the random variables through the conditional distributions of the vertices given the values of its parents. The joint distribution over the set of vertices V in a DAG can be expressed as

$$f(\mathbf{x}) = \prod_{v \in V} f(x_v | x_{\mathrm{pa}(v)}),$$

where $\mathrm{pa}(v)$ is the set of vertices which are the parents of the node v.

A Markov network can be constructed from a DAG by removing the directions on the edges and adding edges between the parents of each vertex. Such a graph allows for the use of a wide range of tools available for Markov networks, see Ripley (1996) [58]. Obviously, a Markov network which is a tree is a Markov tree.

To see how a distribution can be modeled by a Markov tree, Bedford & Cooke (2002) [10] define the following.

Definition 5.1 *(Bivariate tree specification)* $(\underline{F}, T, B)$ *is an n-dimensional bivariate tree specification if*

1. $\underline{F} = (F_1, ..., F_n)$ *is a vector of one-dimensional distribution functions,*
2. T *is an undirected acyclic graph with nodes* $N = \{1, ..., n\}$ *and edges* E*,*
3. $B = \{B(i,j) | \{i,j\} \in E\}$*, where* $B(i,j)$ *is a subset of the class of copula distribution functions.*

Definition 5.2 *(Tree dependence)*

1. *A multivariate probability distribution* G *on* $\Re^n$ *realizes a biviriate tree specification* $(\underline{F}, T, B)$ *if the marginal distributions of* G *are* $F_i, i = 1, ..., n$ *and if for any* $\{i,j\} \in E$ *the bivariate copula* C_{ij} *of* G *is an element of* $B(i,j)$*.*
2. G *has tree dependence of order* M *for* T *if* $\{i,k_1\}, ..., \{k_m, j\} \in E$ *implies that* X_i *and* X_j *are conditionally independent given any* M *of* $k_l, 1 \leq l \leq m$*; and if* X_i *and* X_j *are independent when there are no such* $k_1, ..., k_m$ $(i,j \in N)$*.*
3. G *has Markov tree dependence for* T *if* G *has tree dependence order* M *for every* $M > 0$*.*

An attraction of specifying a multivariate distribution through a Markov tree can be seen from the following theorem, the proof of which can be found in Cooke (1997) [21].

Theorem 5.3 *Let* $(\underline{F}, T, B)$ *be an n-dimensional bivariate tree specification that specifies the marginal distributions* $f_i, i = 1, ..., n$*, and the bivariate densities* $f_{ij}, \{i,j\} \in E$*, the set of edges of* T*. Then there is a unique density* f *on* $\Re^n$ *with marginals* $f_1, ..., f_n$ *and bivariate marginals* $f_{ij}, \{i,j\} \in E$ *such that* f *has Markov tree dependence described by* T*. The density of* f *is given by*

$$f(x_1, ..., x_n) = \frac{\prod_{\{i,j\} \in E} f_{ij}(x_i, x_j)}{\prod_{i \in N} (f_i(x_i))^{\deg(i)-1}}, \tag{5.1}$$

where $\deg(i)$ *denotes the degree of node* i*, that is, the number of neighbours of* i *in the tree* T*.*

The theorem provides a convenient factorization of the probability density for a distribution modeled as a bivariate Markov tree.

As a motivation for the introduction of vines, consider the following simple example, as given in Bedford & Cooke (2001) [9]. We want to specify the joint distribution of three random variables X_1, X_2 and X_3 with uniform marginals. Figure 5.1 shows examples of (a) a belief network, (b) a Markov tree, and (c) a vine.

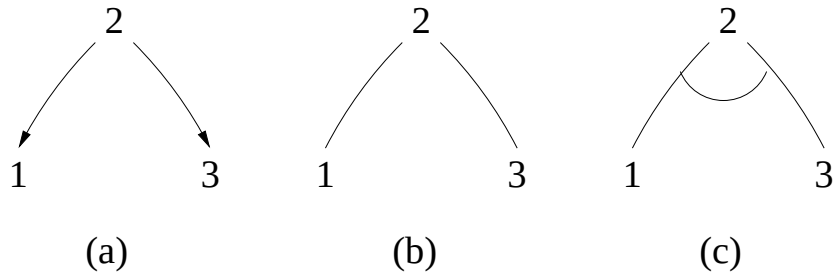


Figure 5.1: A belief network, a Markov tree and a vine

The configuration of the belief network can be hard to do, as it involves specifying conditional distributions in such a way that the remaining marginals are uniform. Besides, the model specifies conditional independence between variables 1 and 3 given the value of variable 2, which may not be what we are looking for. The Markov tree allows the use of copulae as indicated above, making it more convenient for modeling with given marginals. However, the conditional independence assumption remains. Finally, the vine presents a modeling framework where copula modeling is natural while at the same time variables 1 and 3 can be assigned conditional dependence given the value of variable 2, allowing for more flexibility in the choice of dependence structure.

An informal definition of a vine and a regular vine can be found in Kurowicka & Cooke (2002) [45]: They describe a *vine* on n variables as a nested set of trees, where the edges of tree j are the nodes of tree $j+1$, $j = 1, ..., n-2$, and each tree has the maximum number of edges. Further, a *regular vine* on n variables is a vine in which two edges in tree j are joined by an edge in tree $j+1$ only if they share a common node, and this applies for $j = 1, ..., n-2$. Consequently, a regular vine contains $n(n-1)/2$ edges.

The following, more formal, definition can be found in Bedford & Cooke (2002) [10].

Definition 5.4 *(Vine, regular vine)* $\mathcal{V}$ *is a vine on* n *elements if*

1. $\mathcal{V} = (T_1, ..., T_n)$.
2. T_1 *is an undirected acyclic graph with nodes* $N_1 = (1, ..., n)$ *and a set of edges denoted* E_1.
3. *For* $i = 2, ..., m$, T_i *is an undirected acyclic graph with nodes* $N_i \subset N_1 \cup E_1 \cup E_2 \cup \cdots \cup E_{i-1}$ *and edge set* E_i.

$\mathcal{V}$ *is a regular vine on* n *elements if*

1. $m = n$.

2. *T_i is an undirected tree with edge set E_i and node set $N_i = E_{i-1}$ with $\#N_i = n - (i-1)$ for $i = 1, ..., n$, where $\#N_i$ is the cardinality of the set N_i.*

3. *The proximity condition holds: for $i = 2, ..., n-1$, if $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are two nodes in N_i connected by an edge $e \in E_i$ (recall $a_1, a_2, b_1, b_2 \in N_{i-1}$), then $\#a \cap b = 1$.*

The focus here will be on regular vines. Figures 5.2 and 5.3 show two simple representations of regular vines on four variables.

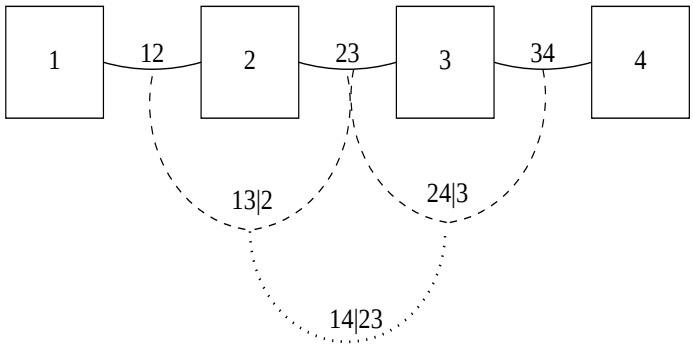


Figure 5.2: The d-vine on four variables

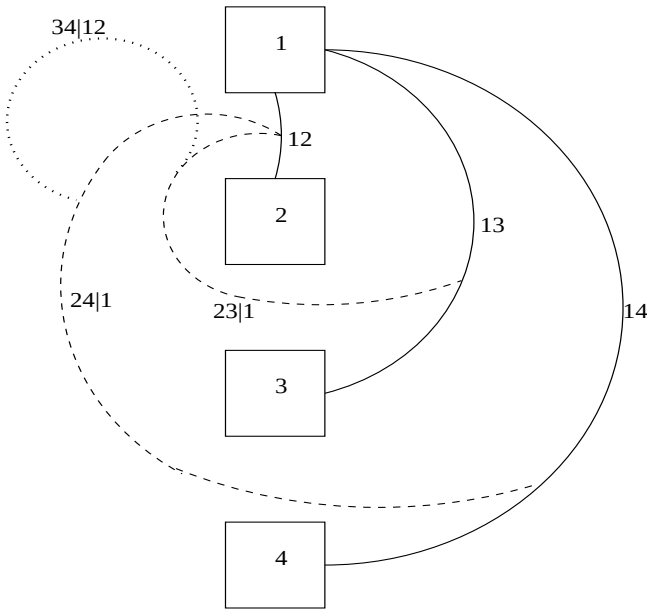


Figure 5.3: The c-vine on four variables

The vine in figure 5.2 is known as a *standard vine*, or *d-vine*, while the one in figure 5.3 is known as a *canonical vine*, or *c-vine*. Note that the structure imposed by the vine depends on the permutation of the variables. Now, the

conditional bivariate distributions of the vines can be expressed through the bivariate version of (4.4), that is,

$$c(F_1(x), F_2(y)) = \frac{f_{12}(x,y)}{f_1(x)f_2(y)}. \tag{5.2}$$

The following example shows how to use this to obtain an expression for the density of the d-vine in figure 5.2.

Example 5.5 (Joint density of the d-vine) The joint density of the 4-variate d-vine can be expressed as

$$\begin{aligned}
f_{1234} &= f_{14|23} f_{23} \\
&= c_{14|23}(F_{1|23}, F_{4|23}) f_{1|23} f_{4|23} f_{23} \\
&= c_{14|23}(F_{1|23}, F_{4|23}) \frac{f_{123} f_{234}}{f_{23}} \\
&= c_{14|23}(F_{1|23}, F_{4|23}) \frac{f_{13|2} f_2 f_{24|3} f_3}{f_{23}} \\
&= c_{14|23}(F_{1|23}, F_{4|23}) c_{13|2}(F_{1|2}, F_{3|2}) c_{24|3}(F_{2|3}, F_{4|3}) \frac{f_{1|2} f_{3|2} f_2 f_{2|3} f_{4|3} f_3}{f_{23}} \\
&= c_{14|23}(F_{1|23}, F_{4|23}) c_{13|2}(F_{1|2}, F_{3|2}) c_{24|3}(F_{2|3}, F_{4|3}) \frac{f_{12} f_{23} f_{34}}{f_2 f_3}.
\end{aligned} \tag{5.3}$$

We see that the final factor in the last line of (5.3) is equal to what we would get using (5.1) for the first-level tree. Thus, Markov tree dependence could be obtained as a special case here by modeling the conditional dependencies by the product copula, i.e. assuming conditional independence.

Bedford & Cooke (2001) [9] generalize the density expression above for regular vines. We will adopt their notation of denoting the *conditioning set* by D_e, that is, the variables conditioned on for an edge labeled $ij|D_e$. The distribution and density of the copula corresponding to this edge are then $C_{ij|D_e}$ and $c_{ij|D_e}$, and the vine dependent density is given in the following theorem.

Theorem 5.6 *Let $\mathcal{V} = (T_1, ..., T_n)$ be a regular vine on n elements. Given F_i and $C_{ij|D_e}$ there is a unique vine dependent distribution with density given by*

$$f_{1\cdots n} = \left(\prod_{m=2}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}) \right) \frac{\prod_{(i,j)\in E_1} f_{ij}}{\prod_{i \in N_1} (f_i)^{\deg(i)-1}}, \tag{5.4}$$

where e is an edge with label $ij|D_e$.

Thus the density of the regular vine dependent distribution can be expressed as the density of the first-level tree as if assumed markov tree dependent, multiplied by the conditional copula densities of the higher-level trees. The density can be simplified a bit by applying (5.2) to the f_{ij}, resulting in the following expression:

$$f_{1\cdots n} = f_1 \cdots f_n \prod_{m=1}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(F_{i|D_e}, F_{j|D_e}). \tag{5.5}$$

Example 5.7 (Joint density and conditional densities of the c-vine) Applying (5.5) to the 4-variate c-vine in figure 5.3, we get the following expression:

$$\begin{aligned} f_{1234} &= f_1 f_2 f_3 f_4 \\ &\quad \times c_{34|12}(F_{3|12}, F_{4|12}) c_{23|1}(F_{2|1}, F_{3|1}) c_{24|1}(F_{2|1}, F_{4|1}) \\ &\quad \times c_{12}(F_1, F_2) c_{13}(F_1, F_3) c_{14}(F_1, F_4). \end{aligned}$$

Inspecting this expression and comparing with the standard density decomposition

$$f_{1234} = f_1 f_{2|1} f_{3|21} f_{4|321},$$

we see the following simple expressions for conditional densities:

$$\begin{aligned} f_1 &= f_1, \\ f_{2|1} &= c_{12} f_2, \\ f_{3|21} &= c_{23|1} c_{13} f_3, \\ f_{4|321} &= c_{34|12} c_{24|1} c_{14} f_4. \end{aligned}$$

Bedford & Cooke (2001) [9] report that the d-vine does not have such a simple expression for conditional densities.

5.1 Sampling the regular vines

Bedford & Cooke (2001) [9] give the following general procedure for sampling a d-vine.

Example 5.8 (Sampling the d-vine) For a d-vine such as the one shown in figure 5.2 sampling can be done by the following scheme.

- Sample X_1 from the distribution F_1.
- From C_{12}, F_1 and F_2, determine $F_{2|1}$ and sample X_2 given X_1.
- From C_{12}, F_1 and F_2, determine $F_{1|2}$. From C_{23}, F_2 and F_3 determine $F_{3|2}$. From $C_{13|2}$, $F_{1|2}$ and $F_{3|2}$ determine $F_{3|12}$ and sample X_3 given X_1 and X_2.
- From C_{34}, F_3 and F_4 determine $F_{4|3}$. From C_{23}, F_2 and F_3 determine $F_{4|3}$. From C_{23}, F_2 and F_3 determine $F_{2|3}$. From $C_{24|3}$, $F_{2|3}$ and $F_{4|3}$ determine $F_{4|23}$. From $C_{13|2}$, $F_{1|2}$ and $F_{3|2}$ determine $F_{1|23}$. From $C_{14|23}$, $F_{1|23}$ and $F_{4|23}$ determine $F_{4|123}$ and sample X_4 given X_1, X_2 and X_3.
- ...

A similar procedure can be outlined for the c-vine as follows:

Example 5.9 (Sampling the c-vine) For a c-vine such as the one shown in figure 5.3 sampling can be done by the following scheme.

- Sample X_1 from the distribution F_1.
- From C_{12}, F_1 and F_2, determine $F_{2|1}$ and sample X_2 given X_1.

- From C_{13}, F_1 and F_3 determine $F_{3|1}$. From $C_{23|1}$, $F_{2|1}$ and $F_{3|1}$ determine $F_{3|12}$ and sample X_3 given X_1 and X_2.
- From C_{14}, F_1 and F_4 determine $F_{4|1}$. From $C_{24|1}$, $F_{2|1}$ and $F_{4|1}$ determine $F_{4|12}$. From $C_{34|12}$, $F_{3|12}$ and $F_{4|12}$ determine $F_{4|123}$ and sample X_4 given X_1, X_2 and X_3.
- ...

The examples above show only schematic procedures. For details on the estimation and sampling of the Student's t c-vine, see appendix A.4.

Chapter 6

Application to fixed income derivative portfolios

This chapter will deal with applying techniques from previous chapters to the forward rates of the Norwegian fixed income market. The reason for modeling forward rates is that we want to calculate market risk for portfolios of fixed income derivatives. Market practice is to price instruments such as caps, floors or european swaptions through variations of *Black's formula*, also known as *Black '76*. This formula was originally introduced in Black (1976) [14] for pricing commodity options. These formulae use forward rates as inputs, and can be shown (see e.g. Björk (2004) [13]) to have a sound theoretical justification concerning application to such instruments. See appendix A.1 for the price formulae of a few common fixed income instruments. Also, a possible extension to include more complex interest rate derivative products in the portfolios motivates the modeling of forward rates. Some of the current state-of-the-art techniques to price such instruments are based on *LIBOR market models*, which model a discrete set of forward rates. See Rebonato (2002) [55] for a thorough discussion on the use of these models.

An important distinction should be noted on the approach taken here, concerning short-term market risk, as compared to models for various long-term risks associated with fixed income contracts. As noted in Jorion (2000) [41]; if the underlying process is a simple diffusion, the innovation component will severely dominate the drift over short time horizons. The drift term scales with Δt, while the innovation term scales with $\sqrt{\Delta t}$, so over a time horizon of, say, one day, the drift coefficient would need to be roughly 20 times as large as the diffusion coefficient to make much of an impact. Models applying over longer time periods would need some appropriate way to reproduce the real drift between rates of different maturities, see e.g. Rebonato et al. (2005) [56]. Such models might be necessary for instance for assessment of counterparty credit risk, hedging performance of models or of investment strategies. Here, though, we are concerned with market risk over a short time period and hence rather put the effort into modeling movements in the innovations.

The use of heavy-tailed models for the short-term real dynamics of returns

combined with Black-type pricing models with their normality assumptions is obviously inconsistent. However, Glasserman, Heidelberger & Shahabuddin (2002) [34] report that option pricing formulae are routinely used this way in practice. We will not be concerned with this problem, rather leaving the question of suitability of the model to the backtesting results.

The data set used for testing the different models is daily NIBOR data from 01.10.98 to 18.08.05, a total of 1750 points[1]. Rates of eight different maturities were used: 3M, 6M, 9M, 1Y, 2Y, 3Y, 5Y and 10Y. This means that when transformed to forward rates, the first rate is the one applying from 0-3M, and the last is the one applying from 5Y-10Y.

Figures 6.1 and 6.2 outline the steps involved in calculating VaR and ES one time period ahead. For the backtesting, the results of which are described in section 6.4, this procedure is repeated for each of 1500 data points and compared to realized changes in portfolio value to evaluate of the appropriateness of the different model specifications.

6.1 The volatility structure and conditional distributions of forward rates

As discussed in chapter 3, we want to specify a volatility strucure accompanied by a conditional distribution which provides a reasonable fit to the actual marginal forward rates. The motivation for this step of the modeling is the autocorrelation which is evident in the time series of the rate changes.

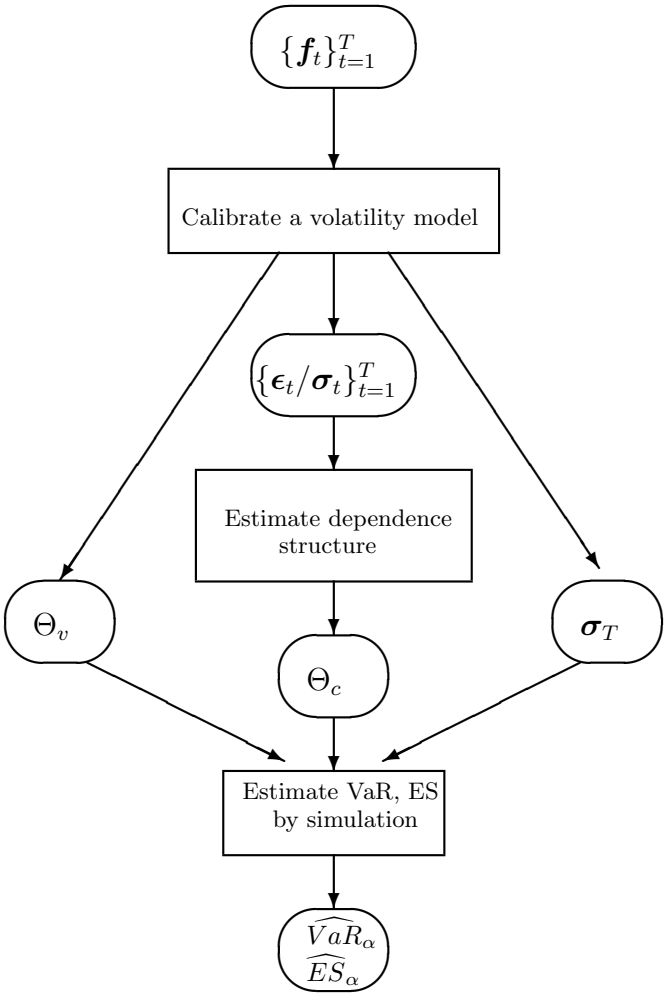
First, we will take a look at the standard AR(1)-GARCH(1,1)-model and the three suggestions for conditional distributions from chapter 3. As an example, consider the entire series of the 2-3 year rate, that is, 1750 observations of the rate applying from two years ahead to three years ahead.

For purposes of comparison, figure 6.3 shows a normal quantile-quantile plot and the autocorrelation function of the absolute values of the changes in this rate. Heavy tails are evident, and also significant autocorrelation at lags up to and above 30.

Turning to our models, we start off with conditionally normal returns. Figure 6.4 shows a normal q-q plot and the autocorrelation function of the absolute values of the standardized residuals ϵ_t/σ_t. As could be expected, both tails of the distribution of the standardized residuals are heavier than those of the normal distribution. The figure on the right shows there is some autocorrelation left at lag one with this model, otherwise no significance is evident.

Figure 6.5 shows the results of fitting the GARCH-model with conditionally Student's t distributed returns. The quantile-quantile plot here shows the fit of the standardized residuals against a Student's t distribution with the estimated degrees of freedom. The residuals appear to be fitted quite well by this

[1]The data was supplied by DnBNOR, through thesis supervisor Jacob Laading



Notation:

$\{\mathbf{f}_t\}_{t=1}^T$, $\mathbf{f}_t = (f_{t1}, ..., f_{tn})$:	n historical forward rates from time $t = 1$ to $t = T$
$\Theta_v = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$:	n vectors of volatility model parameters
$\{\boldsymbol{\epsilon}_t/\boldsymbol{\sigma}_t\}_{t=1}^T$, $\boldsymbol{\epsilon}_t/\boldsymbol{\sigma}_t = (\epsilon_{t1}/\sigma_{t1}, ..., \epsilon_{tn}/\sigma_{tn})$:	Standardized residuals
$\boldsymbol{\sigma}_T = (\sigma_{1,T}, ..., \sigma_{n,T})^\top$:	Current volatility
Θ_c:	Copula parameters

Figure 6.1: Chart of steps involved in obtaining estimates of VaR and ES

Step 1: Estimate the volatility structure

- **Input**: $\{\mathbf{f}_t\}_{t=1}^T$, $\mathbf{f}_t = (f_{t1}, ..., f_{tn})$: n historical forward rates from time $t = 1$ to $t = T$, with time resolution $\Delta t = t_{i+1} - t_i$.
- Calibrate a volatility model to each of the n time series of forward rates.
- **Output**: Matrix $\Theta_v = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ of n vectors of volatility model parameters; Standardized residuals $\{\boldsymbol{\epsilon}_t/\boldsymbol{\sigma}_t\}_{t=1}^T$, $\boldsymbol{\epsilon}_t/\boldsymbol{\sigma}_t = (\epsilon_{t1}/\sigma_{t1}, ..., \epsilon_{tn}/\sigma_{tn})$; Current volatility $\boldsymbol{\sigma}_T = (\sigma_{1,T}, ..., \sigma_{n,T})^\top$.

↓

Step 2: Estimate the dependence structure

- **Input**: Standardized residuals $\{\boldsymbol{\epsilon}_t/\boldsymbol{\sigma}_t\}_{t=1}^T$.
- Calibrate a chosen copula to the standardized residuals.
- **Output**: Copula parameters Θ_c.

↓

Step 3: Simulation

- **Input**: Copula parameters Θ_c; volatility model parameters Θ_v; current volatility $\boldsymbol{\sigma}_T$.
- Repeat for $j = 1, ..., N$ iterations
  - Draw n dependent uniform random variables $u_1, ..., u_n$ according to chosen copula with parameters Θ_c.
  - Invert $u_1, ..., u_n$ according to chosen volatility model with parameters Θ_v and current volatility σ_T to obtain estimated forward rates $\hat{\mathbf{f}}^{T+\Delta t} = (\hat{f}_1^{T+\Delta t}, ..., \hat{f}_n^{T+\Delta t})^\top$.
  - Calculate the value of the portfolio PV_j for the rates $\hat{\mathbf{f}}^{T+\Delta t}$.
- **Output**: $\widehat{VaR}_\alpha$, $\widehat{ES}_\alpha$, estimated from the empirical distribution of simulated portfolio values $\mathbf{PV}$.

Figure 6.2: Chart of steps involved in obtaining estimates of VaR and ES

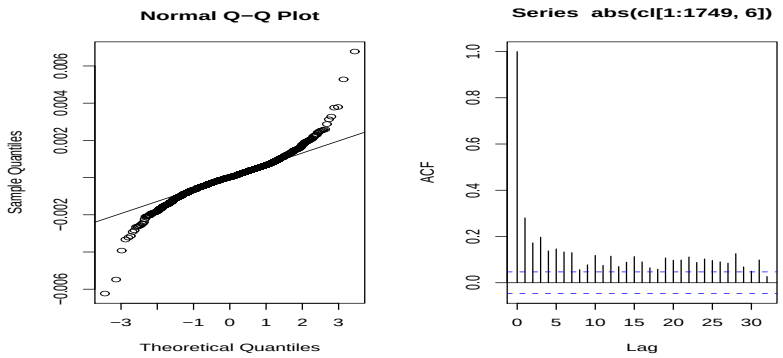


Figure 6.3: Autocorrelation and Q-Q plot for the 2-3 year forward rate changes

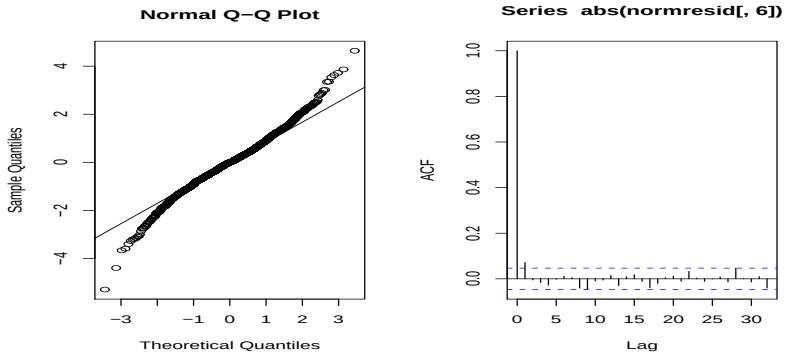


Figure 6.4: Autocorrelation and Q-Q plot for the residuals from the 2-3 year rate assuming conditional normality

t-distribution, though there is some heaviness in the tails that has not been accounted for, particularly in the left tail. The ACF resembles that of the conditionally normal model above.

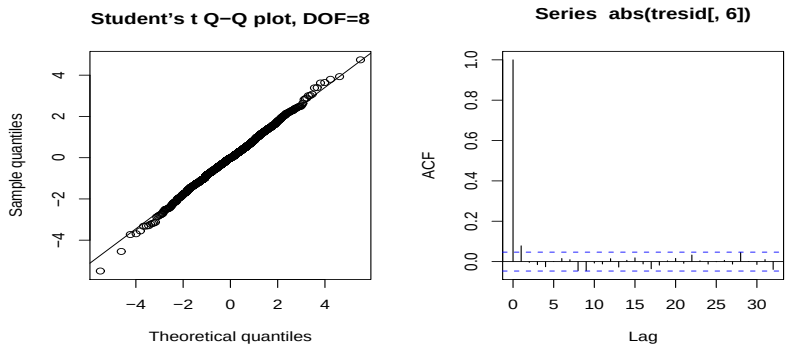


Figure 6.5: Autocorrelation and Q-Q plot for the residuals from the 2-3 year rate assuming a conditional Student's t distribution

Turning to the conditionally IHS-distributed returns, we take a slightly different approach to examine the quantiles. When estimating a GARCH-model with an IHS conditional distribution, Ramirez (2001) [54] points out that if the distributional assumption is correct, then the transformed residuals

$$v_t = \frac{\sinh^{-1}(\epsilon_t/\sigma_t s + \mu) - \lambda}{\delta} \tag{6.1}$$

should be normally distributed. Accordingly, the left plot in figure 6.6 shows a normal q-q plot of the transformed standardized residuals, which appears to fit very well. Again, the ACF resembles the two cases above.

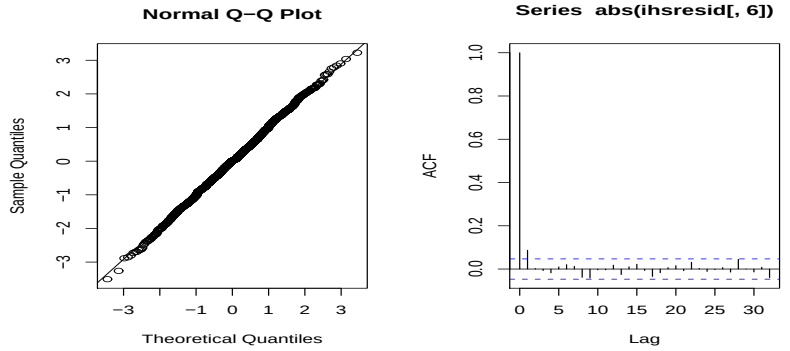


Figure 6.6: Autocorrelation and Q-Q plot for the transformed residuals from the 2-3 year rate assuming a conditional IHS distribution

Note, however, that the above figures describe the fit when the entire data set is used to estimate the model. As model suitability will ultimately be judged

from a risk estimation backtesting procedure, where the model is fitted to the past 250 data points, these results may be misleading.

6.2 The dependence structure of yields vs. forward rates

The underlying factors of fixed-income markets are quoted as *yields*. A yield is the interest rate applying from today's date up until some specified future point in time, known as the *maturity* of the rate. Yields are often defined through the corresponding unit zero-coupon bond price, which is today's price of a bond with payoff 1 at the time of the yield's maturity. Given today's date t_0, a set of maturities $t_1, ..., t_n$ and their corresponding yields $R_1(t_0), ..., R_n(t_0)$, the zero-coupon bond prices $P_1(t_0), ..., P_n(t_0)$ of these maturities are given by

$$P_i(t_0) = \frac{1}{(1 + R_i(t_0))^{(t_i - t_0)}},\ i = 1, ..., n,$$

where a discrete compounding of interest is assumed.

The factors we are interested in modeling here, however, are the *forward rates* of the market, i.e. the rates applying between two future points in time. Given $P_1(t_0), ..., P_n(t_0)$ as above, the set of discrete forward rates are given by

$$f_i(t_0) = \frac{P_i(t_0) - P_{i+1}(t_0)}{1 + [t_{i+1} - t_i]P_{i+1}(t_0)},\ i = 0, ..., n - 1,$$

where $P_0(t_0)$ has been set to 1.

So, what happens to the factors of our market when we convert them from yields to forward rates? The left plot in figure 6.7 shows the 2- and 3-year yields plotted against each other.

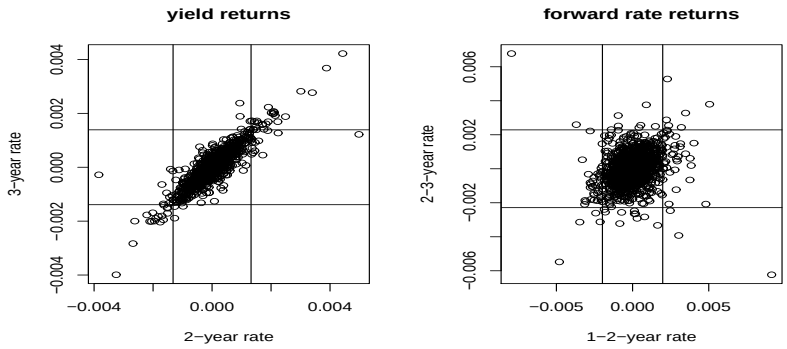


Figure 6.7: Example scatterplots of yields and forward rates

The yield returns appear highly dependent, which should not be a big surprise. The lines in the plot are the marginal 1% and 99% quantiles under

the normality assumption. Frequent and large joint movements outside these bounds may indicate a level of tail dependence incompatible with a normal dependence structure assumption.

The plot on the right-hand side of figure 6.7 shows the corresponding forward rates, namely the rate applying from one to two years and the rate applying from two to three years. Both dependence in general and extreme joint movements are much less evident here. Apart from this, we see that the tails of the marginal distributions, in particular the 1-2-year rate, may be a bit too heavy for the normality assumption, a point we will return to below.

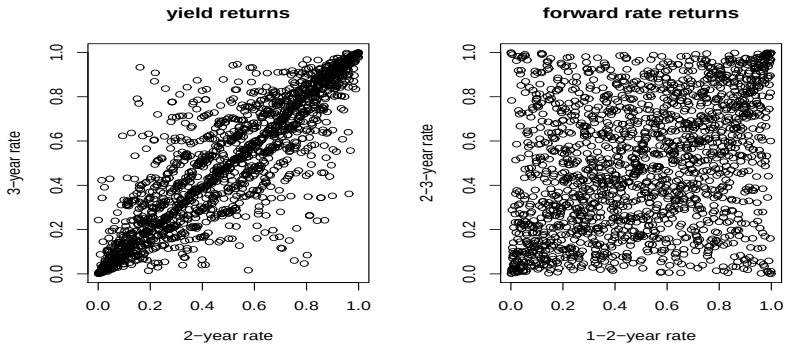


Figure 6.8: Example scatterplots of yields and forward rates when mapped to the unit square

First, have a look at the empirical marginal distributions associated with the same data in figure 6.8. Again, the very high degree of dependency between the yield returns is evident, while the forward rate returns appear much more scattered. For both plots, we see a certain clustering in the lower-left and upper-right corners which might indicate tail dependence, though this is also more obvious for the yield returns. We will not try to estimate the dependence between yields here, but Junker, Szimayer & Wagner (2004) [42] report good results with a transformation of the Frank copula when estimating a form of dependence between monthly 1- and 5-year US yields.

In fact, the strong level of dependence between yield returns has been used to reduce the dimensionality of problems involving a large set of yields, as explored by Litterman & Scheinkman (1991) [49]. The idea of this dimensionality reduction is to take a set of yields and apply some form of factor analysis, such as principal components analysis (PCA). The subject of how to best process the data before applying the PCA is discussed in Lardic, Priaulet & Priaulet (2001) [46], where the authors suggest using the correlation matrix of yield changes. The PCA is done through an eigenvalue decomposition; if the correlation matrix is denoted by R, this decomposition will produce matrices V and Λ such that

$$R = V \Lambda V^{\top},$$

where Λ is a diagonal matrix with the eigenvalues as its elements, and the

columns of V are the corresponding eigenvectors. The first principal component (PC) is then the square root of the largest eigenvalue multiplied by its eigenvector, the second PC is formed the same way from the second largest eigenvalue and its eigenvector and so on. The relative size of each eigenvalue is taken to represent the degree of explanation of that PC, i.e. the degree of the joint movements in the rates explained by this factor. The first three factors are known to exhibit certain qualitative features, labeled *level*, *steepness* and *curvature* by Litterman & Scheinkman (1991) [49]. The level factor is commonly reported to be by far the largest factor. It has the same sign for all maturities, corresponding to a parallell shift in all rates. The steepness factor has different sign for long and short rates, corresponding to a movement where long rates move in the opposite direction of short rates. The curvature factor has the same sign for long and short rates and the opposite sign for intermediate rates, corresponding to a movement where intemediate rates move in the opposite direction of the long and short rates.

Results of such a PCA analysis tend to look a bit different when applied to forward rates, as the first three factors derived from forward rates commonly have smaller degrees of explanation and less obvious qualitative features. Lekkos (2000) [47] suggests that much of the effects found through a factor analysis of yields are due to yields being a form of averages of forward rates. No detailed discussion of this will be given here, we will restrict ourselves to considering figure 6.9. This figure shows the first three factors extracted from the correlation matrix of daily changes in yields and forward rates, respectively.

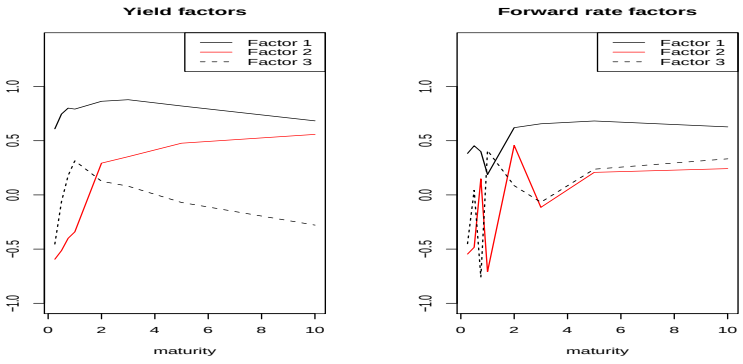


Figure 6.9: Principal components for yields and forward rates

The plot shows how a lot of the structure present in yields disappears when converted to forward rates. The eigenvalues corresponding to the first three factors are also substantially smaller for the forward rates. Still, the level effect is quite pronounced in the forward rates as well as in the yields. In light of the discussion in section 4.2, however, one should be wary of methods expressing dependence solely through linear correlation. Regardless of what explanatory features are found in a factor analysis of the data, there may still remain dependencies which are not expressed through a linear correlation matrix.

6.3 Estimation of the dependence structure of the forward rates

The dependence structure of the forward rates will be estimated from the standardized residuals of the GARCH-estimation described in section 6.1. In other words, a volatility structure is imposed on and calibrated to each marginal forward rate, and the estimation of multivariate dependence is then done on the standardized residuals from these separate univariate procedures. For a similar approach, see Junker & May (2005) [43], who also report that it is common practice in literature to treat the estimated standardized residuals as true observations.

What we are looking for is a model for the dependence that captures as much as possible of the typical co-movements seen in the estimated standardized residuals. It turns out that the dependence between forward rates of different maturities is not entirely trivial. One thing is that, as might be expected, there is a much stronger dependence between adjacent rates than between rates further apart. Also, some rates, to a certain degree, depend negatively on each other. Figure 6.10 shows contourplots of two-dimensional kernel density estimates for a few pairs of residuals mapped to the unit square[2]. The residuals are taken from the estimation of the GARCH model with conditionally Student's t distributed returns, with the estimation done on the entire data set. The left plot shows a clear positive dependence between the rates, the middle plot shows a moderate negative dependence, while the right plot is somewhat harder to interpret.

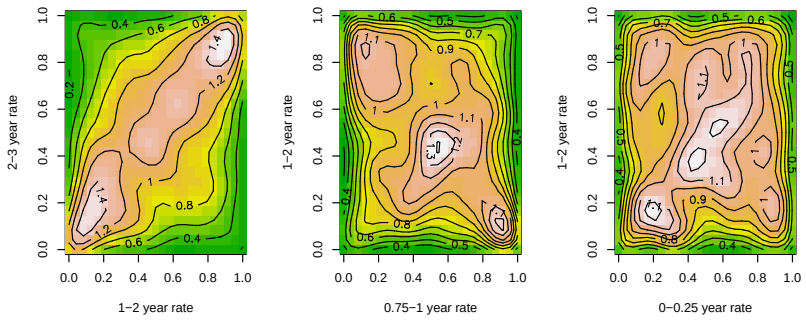


Figure 6.10: Contour plots of various maturity realized residuals mapped to the unit square

Ideally, simulations should reproduce these kinds of co-movements. Figure 6.11 shows corresponding contour plots constructed from simulated values from the Student's t copula. Again, the GARCH estimation is done with a conditional Student's t distribution, and the copula parameter estimation is performed on

[2]Gaussian kernel densities are not ideal for copula density estimation, as it will produce a downslope in the tails, which is spurious. However, keeping this in mind, such estimates should still give a reasonable impression of what the actual density looks like.

the entire data set. Similarly, figure 6.12 is constructed from values simulated with the Student's t c-vine.

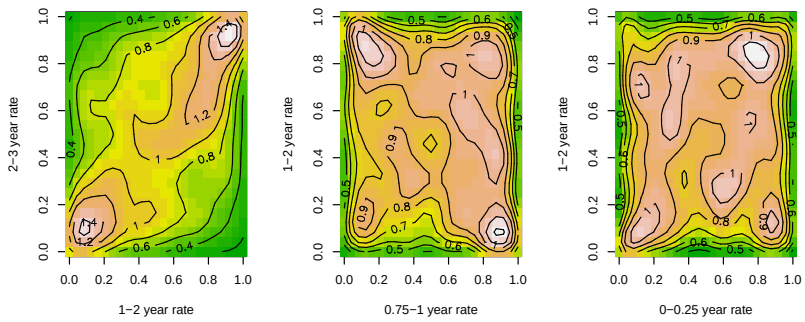


Figure 6.11: Contour plots of various maturity simulated values from the Student's t copula

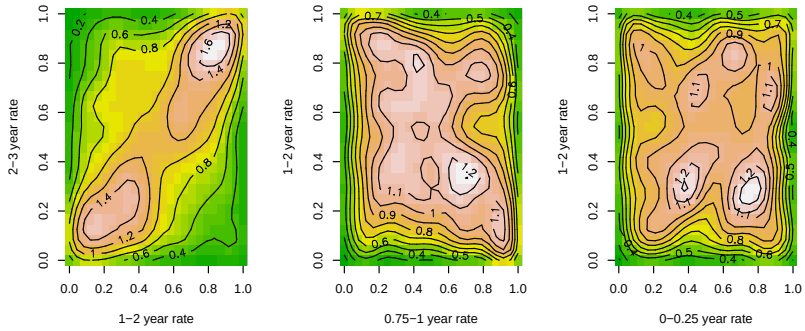


Figure 6.12: Contour plots of various maturity simulated values from the Student's t c-vine

We see that the general patterns from the real data are reproduced by both methods, but neither shows a perfect fit. It appears that the vine does a better job at not putting too much of the density in the tails.

Figure 6.13 shows the same type of plot made from values simulated from the Gaussian copula. Again, the general patterns are more or less represented, but the tails appear a bit on the light side compared to the realized values in figure 6.10.

Table 6.1 shows the results of applying the GOF test discussed in section 4.4.1 to the entire data set of forward rate changes[3]. The test was run with no

[3]The `Copulae R` library of Berg & Bakken was used for the test.

tail weight and power tail weight, for the Student's t and Gaussian copulae. As noted in Demarta & McNeil (2004) [22], the Gaussian copula can be thought of as a special case of the Student's t copula as $\nu \to \infty$, so in a sense we are testing whether the degrees of freedom for the Student's t copula should approach infinity. 5000 simulations were used for each test. The results show clearly that the Gaussian copula is rejected in any case. The Student's t copula appears to give a reasonable fit when the test emphasizes the tails. When no particular region is weighted, both the Student's t copula and the Gaussian copula are rejected at the 5% level, though the Student's t copula attains a higher p-value. This may suggest that the Student's t copula does not capture the entire dependence structure of the distribution, but does a reasonable job of fitting the tails, which for many applications is the most critical part. Applying the same tests to standardized residuals from the volatility model estimation yields similar results.

	No tail weight	Power tail weight, $\alpha = 2$
Student's t copula	0.013	0.404
Gaussian copula	≈ 0	≈ 0

Table 6.1: P-values of the GOF test of Student's t vs. Gaussian copula

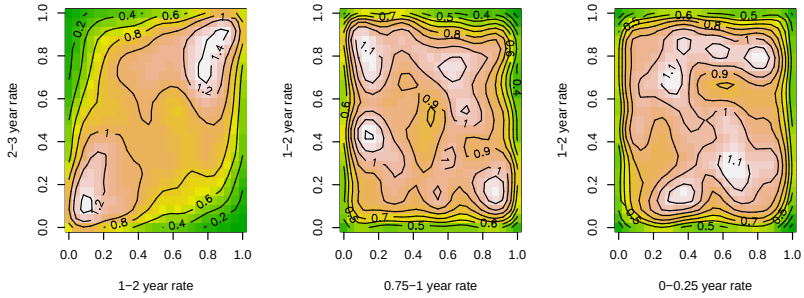


Figure 6.13: Contour plots of various maturity simulated values from the Gaussian copula

As an experiment, the Clayton and Frank copulae were attempted fitted to the residuals as well, with a specification of the form (4.13). The results, however, were not encouraging. The Frank copula tended to underestimate both tails, whereas the Clayton copula appeared to overestimate the tail associated with negative rate returns. For the Clayton copula this should not be a big surprise, as the data do not in general show asymmetric lower tail dependece. In any case, a specification of the type (4.13) amounts to estimating the entire dependence structure of an eight-dimensional distribution with a single scalar parameter, which is a bit optimistic.

6.4 Results of the backtesting procedure

As mentioned in the beginning of this chapter, 1750 days of data were used for testing the models. At each time point, each model was estimated from the past 250 days of data, corresponding to approximately one trading year. Accordingly, 1500 estimates were obtained. In each time step, 10000 iterations were used in the simulation. The estimated one-day VaR and ES estimates for six different quantiles were then compared to the realized daily portfolio return at each time point. For the VaR estimates, two statistics were recorded. One is the number of violations x^α. A good VaR estimate should yield a percentage of violations x^α/N, where N is the sample size, close to the corresponding α-level; i.e. 5% for $\alpha = 0.05$ and $\alpha = 0.95$ and so on. The other is the P-value for the likelihood ratio test statistic used in Aas & Haff (2005) [4], attributed to Kupiec. Under H_0, this test statistic is given by

$$2\log\left(\left(\frac{x^\alpha}{N}\right)^{x^\alpha}\left(1-\frac{x^\alpha}{N}\right)^{N-x^\alpha}\right) - 2\log\left(\alpha^{x^\alpha}(1-\alpha)^{N-x^\alpha}\right),$$

asymptotically distributed as $\chi^2(1)$.

For the ES estimates, a measure suggested by Embrechts, Kaufmann & Patie (2005) [27] is used. In the notation of Aas & Haff (2005) [4], it is given by

$$D^\alpha = \frac{|D_1^\alpha| + |D_2^\alpha|}{2}.$$

D_1^α is the standard ES backtest measure

$$D_1^\alpha = \frac{1}{x^\alpha}\sum_{t\in\kappa^\alpha}(R_t - \widehat{ES}_t^\alpha),$$

where κ^α is the set of points where a VaR violation occurs and R_t is the time t portfolio return. D_1^α is obviously strongly dependent on the VaR estimates. To obtain an ES measure which makes sense on its own, the penalty term D_2^α is introduced. D_2^α is defined as follows. Let $\delta_t^\alpha = R_t - \widehat{ES}_t^\alpha$ and denote by τ^α the set of time points for which δ_t^α is less than (or, for short positions in the same portfolio, greater than) its empirical α-quantile. If y^α is the number of times this happens, then

$$D_2^\alpha = \frac{1}{y^\alpha}\sum_{t\in\tau^\alpha}\delta_t^\alpha.$$

Good estimates of expected shortfall should yield low D^α-values.

Two different portfolios are considered. We would like to have one portfolio with a sufficiently nonlinear value function, to see how the risk estimates handle optionality. A portfolio P_1 consisting of a ten-year cap and a ten-year floor, both struck at 5% is chosen for this purpose. Further, for comparison purposes, a portfolio P_2 consisting of a single ten-year swap with the fixed rate set at 5% is also considered.

The tables show that none of the combinations of copula and volatility specification provide a good fit for all of the quantiles considered. Looking at the

		Gaussian copula					
		GARCH			GJR-GARCH		
	Quantile	N	T	IHS	N	T	IHS
P_1	5%	3.4%	4.07%	3.47%	3.6%	4.13%	3.87%
	95%	4.6%	4.67%	4.27%	3.93%	4.6%	4.6%
	1%	0.867%	0.533%	0.533%	1.13%	0.733%	0.333%
	99%	1.33%	0.733%	1.07%	0.933%	0.8%	1.13%
	0.5%	0.667%	0.2%	0.2%	0.6%	0.267%	0.2%
	99.5%	0.867%	0.467%	0.467%	0.667%	0.333%	0.6%
P_2	5%	5.47%	5.67%	5.2%	4.73%	4.8%	5.27%
	95%	3.87%	4.13%	4.2%	3.6%	4.53%	4.8%
	1%	1.13%	0.6%	0.667%	0.733%	0.733%	0.6%
	99%	1%	0.267%	0.467%	0.933%	0.6%	0.2%
	0.5%	0.667%	0.267%	0.333%	0.467%	0.267%	0.4%
	99.5%	0.533%	0.133%	0.0667%	0.533%	0.133%	0.2%

Table 6.2: Percentage violations for the Gaussian copula under different volatility assumptions

		Gaussian copula					
		GARCH			GJR-GARCH		
	Quantile	N	T	IHS	N	T	IHS
P_1	5%	0.00261	0.0869	0.00401	0.00897	0.113	0.0363
	95%	0.471	0.549	0.182	0.0493	0.471	0.471
	1%	0.595	0.0462	0.0462	0.611	0.276	0.00258
	99%	0.217	0.276	0.797	0.793	0.42	0.611
	0.5%	0.384	0.0608	0.0608	0.595	0.159	0.0608
	99.5%	0.0684	0.853	0.853	0.384	0.33	0.595
P_2	5%	0.414	0.246	0.724	0.633	0.721	0.638
	95%	0.0363	0.113	0.144	0.00897	0.4	0.721
	1%	0.611	0.0926	0.167	0.276	0.276	0.0926
	99%	1	0.000693	0.0205	0.793	0.0926	0.000145
	0.5%	0.384	0.159	0.33	0.853	0.159	0.569
	99.5%	0.856	0.0166	0.0027	0.856	0.0166	0.0608

Table 6.3: Kupiec test P-values for the Gaussian copula under different volatility assumptions

		Gaussian copula					
		GARCH			GJR-GARCH		
	Quantile	N	T	IHS	N	T	IHS
P_1	5%	1.06	1.22	1.26	1.05	1.18	1.28
	95%	1.03	1.1	1.08	1.02	1.1	1.04
	1%	0.0527	0.0989	0.0826	0.0209	0.107	0.0575
	99%	1.59	1.95	2.09	1.57	1.96	1.81
	0.5%	2.06	6.28	5.57	2.11	6.15	5.09
	99.5%	1.84	2.45	2.65	1.78	2.44	2.27
P_2	5%	0.64	0.696	0.667	0.637	0.68	0.657
	95%	0.693	0.946	0.807	0.705	0.756	0.83
	1%	0.0247	0.0812	0.0163	0.0102	0.069	0.0223
	99%	1.1	2.24	1.83	1.12	1.8	1.76
	0.5%	1.26	1.87	1.51	1.27	1.66	1.52
	99.5%	1.3	4.69	3.48	1.35	3.93	2.6

Table 6.4: D^{α}-values for the Gaussian copula under different volatility assumptions

		Student's t copula						Student's t c-vine
		GARCH			GJR-GARCH			GARCH
	Quant.	N	T	IHS	N	T	IHS	T
P_1	5%	4.2%	4.27%	4.73%	3.53%	4.8%	4.2%	4.13%
	95%	5%	4.67%	5.13%	4.4%	4.67%	5.4%	4.4%
	1%	1%	0.667%	0.733%	1%	1%	0.533%	0.8%
	99%	0.933%	0.6%	0.733%	0.867%	0.733%	0.933%	0.6%
	0.5%	0.533%	0.333%	0.267%	0.467%	0.267%	0.267%	0.267%
	99.5%	0.667%	0.267%	0.533%	0.533%	0.4%	0.533%	0.267%
P_2	5%	4.87%	5.53%	5.73%	5.13%	4.87%	5.4%	5.2%
	95%	3.87%	4.67%	4.93%	3.33%	5.07%	4.6%	4.4%
	1%	0.933%	0.4%	0.6%	0.867%	0.467%	0.6%	0.4%
	99%	0.8%	0.467%	0.4%	1%	0.533%	0.4%	0.533%
	0.5%	0.4%	0.133%	0.333%	0.533%	0.333%	0.333%	0.2%
	99.5%	0.533%	0.267%	0.133%	0.333%	0.2%	0.2%	0.0667%

Table 6.5: Percentage violations for the Student's t copula under different volatility assumptions, as well as the Student's t c-vine

		Student's t copula						Student's t c-vine
		GARCH			GJR-GARCH			GARCH
	Quant.	N	T	IHS	N	T	IHS	T
P_1	5%	0.144	0.182	0.633	0.00605	0.721	0.144	0.113
	95%	1	0.549	0.813	0.277	0.549	0.483	0.277
	1%	1	0.167	0.276	1	1	0.0462	0.42
	99%	0.793	0.0926	0.276	0.595	0.276	0.793	0.0926
	0.5%	0.856	0.33	0.159	0.853	0.159	0.159	0.159
	99.5%	0.384	0.159	0.856	0.856	0.569	0.856	0.159
P_2	5%	0.812	0.351	0.202	0.813	0.812	0.483	0.724
	95%	0.0363	0.549	0.905	0.00166	0.906	0.471	0.277
	1%	0.793	0.00789	0.0926	0.595	0.0205	0.0926	0.00789
	99%	0.42	0.0205	0.00789	1	0.0462	0.00789	0.0462
	0.5%	0.569	0.0166	0.33	0.856	0.33	0.33	0.0608
	99.5%	0.856	0.159	0.0166	0.33	0.0608	0.0608	0.0027

Table 6.6: Kupiec test P-values for the Student's t copula under different volatility assumptions, as well as the Student's t c-vine

		Student's t copula						Student's t c-vine
		GARCH			GJR-GARCH			GARCH
	Quant.	N	T	IHS	N	T	IHS	T
P_1	5%	1.02	1.28	1.28	1.04	1.2	1.26	1.16
	95%	1.02	1.13	1.09	1.02	1.12	1.05	1.11
	1%	0.0424	0.189	0.0982	0.0216	0.217	0.0796	0.156
	99%	1.64	2.01	2.05	1.58	2.04	1.82	2.05
	0.5%	2.04	4.84	8.63	2.04	4.9	5.3	4.97
	99.5%	1.86	2.66	2.75	1.83	2.72	2.34	2.61
P_2	5%	0.657	0.713	0.676	0.642	0.699	0.658	0.697
	95%	0.686	0.827	0.831	0.705	0.806	0.8	0.775
	1%	0.0293	0.0925	0.0252	0.034	0.0354	0.0348	0.102
	99%	1.1	2.25	2.2	1.12	2.1	1.66	1.66
	0.5%	1.31	1.88	1.48	1.31	1.71	1.46	1.95
	99.5%	1.28	3.19	4.64	1.37	3.36	2.71	2.85

Table 6.7: D^{α}-values for the Student's t copula under different volatility assumptions, as well as the Student's t c-vine

quantiles pairwise and starting with the 5%/95% quantiles, table 6.3 shows that the VaR estimates of the Gaussian copula are frequently rejected at the 0.05 level. A look at the values in table 6.2 shows a tendency to overestimate the VaR at these quantiles. Tables 6.6 and 6.5 show that in terms of number of violations, the Student's t copula fits better here. The Student's t GJR-GARCH appears to give the best estimates when both portfolios are considered. The IHS GARCH gives reasonable results as well. This may suggest that asymmetry should be taken into account in some way, but maybe that it should rather be done in the GARCH equations than in the conditional distribution. Taking a look at the D^{α}-values though, we see that the conditionally normal volatility models give the best ES estimates, regardless of copula choice. This could indicate that though the tails of the Student's t and IHS distributions have the right thickness at the 5%/95% quantiles, they decay too slowly.

Moving further into the tails at the 1%/99% and 0.5%/99.5% quantiles, we see that the conditionally normal volatility models outperform the others. The Kupiec test P-values and percentage violations suggest that in particular the Student's t copula with a conditionally normal GJR-GARCH volatility could be appropriate. The Student's t and IHS distributions appear useless this far into the tails, almost consistently overestimating the risk at both ends of the portfolio value distribution. The D^{α}-values further show that the extreme tails of the marginal distributions are modeled better by the more quickly decaying normal distribution. The discussion in section 6.1 indicated rather that the heavier-tailed conditional distributions appeared to give a better fit. There, however, we looked at the fit when the model was calibrated to the entire data set. The situation appears to shift in favor of the normal distribution when the model is calibrated many times on only the past 250 days of data.

The comparison of contour plots in section 6.3 suggested that when comparing simulated values from the vine to simulated values from the Student's t copula, those from the vine looked more realistic. Thus, we might have expected the vine to provide better risk estimates than the Student's t copula on the conditionally Student's t distributed residuals. Looking at the tables 6.5-6.7, we see that this is not the case, the two models perform quite similarly. This suggests that even if the vine captures more of the dependence structure, it might not be dependence which is important for the risk associated with these two particular portfolios.

Chapter 7

Conclusion

Several points are worth noting in conclusion concerning both the volatility of individual forward rates and the dependence between them.

As for the volatility, the normal distribution seems like a good choice of conditional distribution for sufficiently infrequent events. The tails of the normal distribution appear to decay at an appropriate rate when considering the portfolio value distributions at the 99%/1% quantile and beyond. They do, however, appear too heavy at the 95%/5% quantiles. To obtain a volatility model suitable at all quantiles considered here, one might want a conditional distribution with tails that can be lighter than those of the normal distribution at the 95%/5% quantile, but which decay at a similar rate. The possibility to account for skewness might also be desirable, though perhaps not critical.

Concerning the volatility model itself, the possibility to capture asymmetry appeared favorable in some cases, but overall did not make too much of a difference. The volatility specifications tested here are only two simple choices out of a large number of GARCH-type models suggested for financial data. Bao, Lee & Saltoğlu (2004) [8] list a good number of such models as well as suggestions for conditional distributions. Models that let the volatility depend on the rate level have been reported to give good results for modeling the short rate, and it could be interesting to see whether they could be applied with success to forward rates over different time periods. For more on these models, see Aas (2004b) [2] and the references therein. Finally, the models used here are applied to the univariate time series of individual rates and thus do not take into account possible covariation in volatility across rates. A multivariate GARCH-type model could be applied to capture such effects. For a brief introduction and references to further work on this, see Aas & Dimakos (2004) [3].

Turning to the dependence structure, the Student's t copula seems like the most sensible choice out of the suggestions considered here. As mentioned in section 6.3, the Gaussian copula can be considered a special case of the Student's t copula as $\nu \to \infty$, that is, as the tail dependence disappears. It could have been a reasonable choice in the case of insignificant tail dependence, as it is less computationally demanding to estimate and simulate, but the results indicate that the Student's t copula does indeed give a better fit.

The Student's t c-vine could have been expected to perform better than it does here, at least judging from the inspection of density plots in section 6.3. This vine specification was only run on a single volatility model example due to the computational effort involved, so further testing seems to be in order. Other permutations of the variables could also be considered, or, for that matter, other copulas for the bivariate dependencies. As it appeared to perform similarly to the Student's t copula in the backtest here, the Student's t copula appears as yet the better choice due to significantly lower computational time for estimation and simulation.

The Archimedean copulae used here did not perform well, but the way these models were specified, they could not be expected to. It should not come as a surprise that it is difficult to capture the dependence between eight variables with a single scalar parameter. Specifying these multivariate copulae to allow for different generators, as discussed in section 4.1.2, would be a natural extension to the approach here.

Last, it should be kept in mind that the tests here were restricted to two portfolios, and one may wish to consider other combinations of instruments. Furthermore, the tests were restricted to the one-day horizon; testing on longer horizons would be desirable to obtain more general performance results.

Bibliography

[1] Aas, K. (2004). Modelling the dependence structure of financial assets: A survey of four copulas. Technical report, The Norwegian Computing Centre.

[2] Aas, K. (2004). Modelling the stochastic behaviour of short-term interest rates: A survey. Technical report, The Norwegian Computing Centre.

[3] Aas, K. and Dimakos, X. (2004). Statistical modelling of financial time series: An introduction. Note, The Norwegian Computing Centre.

[4] Aas, K., Haff, I.H. (2005). The generalized hyperbolic skew Student's t distribution. Technical report, The Norwegian Computing Centre.

[5] Abdous, B., Genest, C., and Remillard, B. (2005). Dependence properties of meta-elliptical distributions. *Statistical modeling and analysis for complex data problems* 1-15, Duchesne, P., Remillard, B. eds. GERAD 25th Anniversary Series, vol. 1. Springer, New York.

[6] Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. Abaxbank/Deutsche Bundesbank.

[7] Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1999) Coherent measures of risk. *Mathematical Finance 9 (3)*, 203-228.

[8] Bao, Y., Lee, T. and Saltoğlu, B. (2004). A test for density forecast comparison with applications to risk management. UC Riverside/Marmara University.

[9] Bedford, T. and Cooke, R. (2001) Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of mathematics and artificial intelligence 32*, 245-268.

[10] Bedford, T. and Cooke, R. M. (2002). Vines - A new graphical model for dependent random variables. *The Annals of Statistics 30 (4)*, 1031 - 1068.

[11] Berg, D. and Bakken, H. (2006). Copula goodness-of-fit tests: A comparative study. Working paper, The Norwegian Computing Centre.

[12] Berg, D. and Bakken, H. (2005). A Goodness-of-fit test for copulae based on the probability integral transform. Note, The Norwegian Computing Centre.

[13] Björk, T. (2004). *Arbitrage theory in continuous time*. Oxford University Press.

[14] Black, F. (1976). The pricing of commodity contracts. *Journal of financial economics 3*, 167-179.

[15] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31* 307-327.

[16] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics 69 (3)*, 542-547.

[17] Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G. and Roncalli, T. (2000) Copulas for finance. A reading guide and some applications. Technical report, Credit Lyonnais.

[18] Bradley, B. O. and Taqqu, M. S. (2001). Financial risk and heavy tails. Boston University.

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

[20] Choi, P. (2001) Estimation of Value at Risk using Johnson S_U-normal distribution. Texas A&M University.

[21] Cooke, R. (1997) Markov trees and tree dependent random variables. Delft University of Technology.

[22] Demarta, S., McNeil, A. (2004). The t copula and related copulas. ETH Zürich.

[23] Devroye, L. (1986) *Non-uniform random variate generation*. Springer-Verlag.

[24] Duffie, D. and Pan, J. (1997). An overview of Value-at-Risk. Draft, Stanford University.

[25] Embrechts, P., McNeil, A., Straumann, D. (2002) Correlation and dependency in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and beyond*, ed. Dempster, M. A. H. Cambridge University Press.

[26] Embrechts, P., Lindskog, F. and McNeil, A. (2001) Modelling dependence with copulas and applications to risk management. ETH Zürich.

[27] Embrachts, P., Kaufmann, R., Patie, P. Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications 32*, 61-90.

[28] Fang, H., Fang, K. and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of multivariate analysis 82*, 1-16.

[29] Finger, C. (2006). How historical simulation made me lazy. *RiskMetrics Research monthly, April 2006*

[30] Franke, J., Härdle, W. and Hafner, C. M. (2004). *Statistics of financial markets*. Springer-Verlag.

[31] Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal 2*, 1-25.

[32] Frey, R. and McNeil, A. (2002). VaR and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. ETH Zürich.

[33] Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika 74 (3)*, 549-555.

[34] Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2002) Portfolio Value-at-Risk with heavy-tailed risk factors. *Mathematical Finance 12 (3)*, 239-269.

[35] Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Springer-Verlag.

[36] Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993) On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance 48 (5)*, 1779-1801.

[37] Hult, H. and Lindskog, F. (2001). Multivariate extremes, aggregation and dependence in elliptical distributions. Preprint, RiskLab.

[38] Jaschke, S. (2002) Quantile-VaR is the wrong measure to quantify market risk for regulatory purposes. Weierstrass-institut für angewandte analysis und stochastik.

[39] Joe, H. (1997) *Multivariate models and dependence concepts*. Chapman & Hall.

[40] Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika 36*, 149-176

[41] Jorion, P. (2000) *Value at Risk: The new benchmark for managing financial risk*. McGraw-Hill.

[42] Junker, M., Szimayer, A. and Wagner, N. (2004). Nonlinear term structure dependence: Copula functions, empirics and risk implications. Caesar Financial Engineering / University of Western Australia / Munich University of Technology.

[43] Junker, M., May, A. (2005). Measurement of aggregate risk with copulas. *Econometrics Journal 8*, 428-454.

[44] Kruskal, W.H. (1958). Ordinal measures of association. *Journal of the American Statistical Association 53*, 814-861.

[45] Kurowicka, D. and Cooke, R. M. (2002). The vine copula method for representing high dimensional dependent distributions: Application to continuous belief nets. *Proceedings of the 2002 Winter Simulation Conference.*

[46] Lardic, S., P. Priaulet, and S. Priaulet (2001). PCA of the Yield Curve Dynamics: Questions of Methodologies. *Journal of Bond Trading & Management 1(4)*, 327 - 49.

[47] Lekkos, I. (2000). A critique of factor analysis of interest rates. *Journal of derivatives 8(1)*, 72-83.

[48] Lindskog, F. (2000). Linear correlation estimation. ETH Zürich.

[49] Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of fixed income 1(1)*, 54-61.

[50] Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association 83*, 834-841.

[51] Mashal, R. and Zeevi, A. (2002). Beyond correlation: Extreme co-movements between financial assets. Columbia University.

[52] Melchiori, M. R. (2006). Tools for sampling multivariate Archimedean copulas. Universidad Nacional del Litoral.

[53] Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988). *Numerical recipes in C.* Cambridge University Press.

[54] Ramirez, O. (2001). Autoregressive conditional heteroskedasticity under error-term non-normality. Texas Tech. University.

[55] Rebonato, R. (2002). *Modern pricing of interest-rate derivatives: The LIBOR market model and beyond.* Princeton University Press.

[56] Rebonato, R., Mahal, S., Joshi, M., Buchholz, L. and Nyholm, K. (2005). Evolving yield curves in the real-world measures: A semi-parametric approach. *Journal of Risk 7 (3)*, 29-61.

[57] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23 (3)*, 470-472.

[58] Ripley, B. D. (1996). *Pattern recognition and neural networks.* Cambridge University Press.

[59] Morgan Guaranty Trust Company (1996). RiskMetrics technical document, 4th ed.

[60] Savu, C. and Trede, M. (2004). Goodness-of-fit tests for parametric families of Archimedean copulas. University of Münster.

Appendix A

A.1 Price formulae for common fixed-income instruments

A.1.1 The FRA

The perhaps simplest fixed income derivative instrument is the forward rate agreement (FRA). An FRA is an agreement between two parties at time t to exchange an amount of money at time $T + \tau$ proportional to the difference between a strike K, agreed upon at time t, and the interest rate $R(T, T + \tau)$ that resets at time T for payment at time $T + \tau$. The proportionality factor is given by the notational principal NP times the accrual period τ. Thus the payoff at time $T + \tau$ will be

$$\text{Payoff(FRA)}_{T+\tau} = NP[R(T, T + \tau) - K]\tau. \tag{A.1}$$

Denote by $f(t, T, T + \tau)$ the time t forward rate applying between T and $T + \tau$. The payoff (A.1) can be expressed as

$$\text{Payoff(FRA)}_{T+\tau} = NP[f(T, T, T + \tau) - K]\tau.$$

Further, let $P(t, T)$ denote the time t price of a maturity T unit payoff zero-coupon bond. An arbitrage argument shows (see e.g. Rebonato (2002) [55]) that the time t present value of the FRA is given by

$$\text{PV(FRA)}_t = NP[f(t, T, T + \tau)\tau - K]P(t, T + \tau). \tag{A.2}$$

A.1.2 The cap

A cap is a series of call options on FRAs, with each option on an FRA known as a caplet. The caplet payoff at time $T + \tau$ is thus

$$\text{Payoff(Caplet)}_{T+\tau} = NP[f(T, T, T + \tau) - K]^+\tau. \tag{A.3}$$

where $[\cdot]^+$ denotes $\max(\cdot, 0)$. Black (1976) [14] showed that under the assumption of lognormally distributed forward rates, an arbitrage argument gives the present value of such a contract as

$$\text{PV(Caplet)}_t = NP[f(t, T, T + \tau)\Phi(h_1) - K\Phi(h_2)]P(t, T + \tau)\tau, \tag{A.4}$$

where

$$h_1 = \frac{\log[f(t,T,T+\tau)/K] + \frac{1}{2}\sigma^2_{\text{Black}}(T-t)}{\sigma_{\text{Black}}\sqrt{T-t}} \tag{A.5}$$

$$h_2 = \frac{\log[f(t,T,T+\tau)/K] - \frac{1}{2}\sigma^2_{\text{Black}}(T-t)}{\sigma_{\text{Black}}\sqrt{T-t}}, \tag{A.6}$$

where Φ is the standard normal distribution and σ_{Black} is the *Black volatility*. This is the formula known as *Black '76*. The present value of the cap is simply the sum of the present values of its caplets. The Black volatility is defined as follows. Let σ_i be the instantaneous volatility of the forward rate resetting at time T_i. The Black volatility is related to σ_i through

$$\sigma^2_{\text{Black}}(T_i) = \frac{1}{T_i}\int_0^{T_i} \sigma_i^2(u)\mathrm{d}u. \tag{A.7}$$

In a liquid cap market, standard practice is to calculate the Black volatility from prices of traded caps by inverting Black's formula. Here, however, we stick to the historical interest rate data, and calculate the Black volatility from historical volatilities of forward rates and application of (A.7).

A.1.3 The floor

A floor is a series of put options on FRAs, with each option known as a floorlet. Thus, the payoff of a floorlet is

$$\text{Payoff(Floorlet)}_{T+\tau} = NP[K - f(T,T,T+\tau)]^+\tau. \tag{A.8}$$

Similarly as for caplets, the present value of the floorlet is

$$\text{PV(Floorlet)}_t = NP[K\Phi(-h_2) - f(t,T,T+\tau)\Phi(-h_1)]P(t,T+\tau)\tau \tag{A.9}$$

with h_1 and h_2 defined as above. Again, the present value of the floor is the sum of the present values of its floorlets.

A.1.4 The swap

A swap is an agreement between two parties to exchange the payments represented by interest on a notational principal, on equally spaced prespecified dates. One side pays a fixed rate, the other a floating rate. No principal is exchanged at the maturity of the contract. If the fixed rate is set at K and there are a total of N periods, the annuity made up of the payments on the fixed rate add up to

$$\text{A}_t = NP \cdot K \sum_{i=0}^{N-1} P(t,T_i+\tau)\tau, \tag{A.10}$$

where $T_i + \tau = T_{i+1}$.

The payments on the floating rate, assuming the number of payments are the same as on the fixed rate, become

$$\begin{aligned} \mathrm{FL}_t &= \sum_{i=1}^{N} NPf(t, T_i, T_i + \tau)\tau P(t, T_i + \tau) && \text{(A.11)} \\ &= NP\sum_{i=1}^{N}[P(t, T_i) - P(t, T_i + \tau)] && \text{(A.12)} \\ &= NP[P(t, T_0) - P(t, T_N)]. && \text{(A.13)} \end{aligned}$$

The value of the swap is then $\mathrm{A}_t - \mathrm{FL}_t$ to the receiver of the fixed rate, and $\mathrm{FL}_t - \mathrm{A}_t$ to the receiver of the floating rate.

A.2 Parameter estimation for the volatility specifications

For both GARCH model specifications, in combination with each of the three conditional distributions, the parameters are estimated by a multidimensional optimization of a likelihood. To outline the general procedure, let the volatility specification be given by

$$\begin{aligned} y_t &= x_t^\top b + \epsilon_t \\ \mathrm{E}(\epsilon_t) &= 0 \\ \mathrm{Var}(\epsilon_t) &= \sigma_t^2 \\ \sigma_t^2 &= z_t^\top \theta \end{aligned} \qquad \text{(A.14)}$$

For instance, the standard AR(1)-GARCH(1,1) model (3.2) - (3.5) would have $y_t = r_t$, $x_t = (1, r_{t-1})^\top$, $b = (b_0, (b_1 + 1))^\top$, $z_t = (1, \epsilon_{t-1}^2, \sigma_{t-1}^2)^\top$, and $\theta = (\alpha_0, \alpha_1, \beta_0)^\top$. Also, the conditional distribution employed may contain additional parameters θ^*, i.e. $\theta^* = \{\nu\}$ for the Student's t distribution and $\theta^* = \{\lambda, \delta\}$ for the IHS distribution. Now, we want to maximize the likelihood with respect to all the parameters simultaneously. Algorithm 1 returns the log-likelihood $L_T(b, \theta, \theta^*; \mathbf{x}, \mathbf{y})$ for given parameters b, θ and θ^*, and data $\mathbf{y} = (y_1, ..., y_T)^\top$, $\mathbf{x} = (x_1, ..., x_T)^\top$. The density function $f(\theta, \theta^*; \epsilon, \sigma)$ is the density for the conditional distribution being used.

The specification (A.14) requires starting values for ϵ_0 and σ_0. The value of ϵ_0 is set by including an extra data point (y_0, x_0) used only for this purpose, while σ_0^2 is set to the empirical variance $\widehat{\mathrm{Var}}(\epsilon)$ of the residuals $\epsilon = (\epsilon_1, ..., \epsilon_T)^\top$.

The log-likelihood functions of the three choices of conditional distribution are listed in table A.1

So, the objective is to find the parameters that maximize the function $L_T(b, \theta, \theta^*; \mathbf{x}, \mathbf{y})$ given by algorithm 1, i.e.

$$(b, \theta, \theta^*) = \arg \max_{b, \theta, \theta^* \in \Theta} L_T(b, \theta, \theta^*; \mathbf{x}, \mathbf{y}), \qquad \text{(A.15)}$$

Algorithm 1 $L_T(b, \theta, \theta^*; \mathbf{x}, \mathbf{y})$

Require: Data $y_t, x_t, t = 0, ..., T$, parameters b, θ, θ^*

for $t = 0, ..., T$ **do**
  $\epsilon_t = y_t - x_t^\top b$
end for
$\sigma_0^2 = \widehat{\text{Var}}(\epsilon)$
for $t = 1, ..., T$ **do**
  $\sigma_t^2 = z_t^\top \theta$
end for
$L = \sum_{t=1}^{T} \log f(\theta, \theta^*; \epsilon_t, \sigma_t)$
return L, b, θ, θ^*

Distribution	Log-likelihood L
Gaussian	$-\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2}\right)$
Student's t	$T \cdot \left(\log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log(\pi(\nu-2))\right)$ $-\frac{1}{2}\sum_{t=1}^{T}\left(\log(\sigma_t^2) + (\nu+1)\log(1 + \frac{\epsilon_t^2}{\sigma_t^2(\nu-2)})\right)$
IHS .	$T \cdot \left(-\frac{1}{2}\log(2\pi) + \log(s) - \log(\delta)\right)$ $-\frac{1}{2}\sum_{t=1}^{T}\left(\log(\sigma_t^2) - \log(R^2+1) - \frac{(\sinh^{-1}(R)-\lambda)^2}{\delta^2}\right)$ where $R = \frac{\epsilon_t}{\sigma_t s + \mu}$, $s = \sqrt{\frac{1}{2}(\omega-1)(\omega\cosh(2\lambda)+1)}$, $\mu = \sqrt{\omega}\sinh(\lambda)$, and $\omega = \exp(\delta^2)$

Table A.1: Likelihood functions for the three choices of conditional distributions

with Θ being the parameter space. The method used here is the `amoeba` method from Press et al. (1988) [53], which is an implementation of the *downhill simplex* algorithm. The downhill simplex is not necessarily the fastest method, but it is easy to implement, robust, and requires only function evaluations, no derivatives are needed. Positivity constraints for the parameters θ were imposed by iterating over values $\tilde{\theta}_i$ and setting $\theta = \tilde{\theta}_i^2$.

A.3 Parameter estimation for the copulae

Three of the copulae implemented here require a univariate numerical maximization of a likelihood to estimate a parameter. These are the Student's t copula, for which the degrees of freedom ν must be estimated in this fashion, and the Clayton and Frank copulae, for which the single parameter δ must be estimated. Estimation of the correlation matrix of the Gaussian copula and the scale matrix of the Student's t copula do not require such a procedure, as they can be estimated through the methods described in section 4.3.

A.3.1 The likelihood for the degrees of freedom for the Student's t copula

From (4.11) and (4.29) we see that the maximization problem for the Student's t copula becomes

$$\hat{\nu} = \arg \max_{\nu \in (2,\infty]} \left[\sum_{t=1}^{T} \log c(\mathbf{u}_t; \nu, \hat{R}) \right] \tag{A.16}$$

where the likelihood $\sum_{t=1}^{T} \log c(\mathbf{u}_t; \nu, \hat{R})$ is given by

$$\begin{aligned}
&\sum_{t=1}^{T} \log c(\mathbf{u}_t; \nu, \hat{R}) = \\
&\quad T \cdot \left(\log \Gamma \left(\frac{\nu+1}{2} \right) + (n-1) \log \Gamma \left(\frac{\nu}{2} \right) - \frac{1}{2} \log |R| - n \log \Gamma \left(\frac{\nu+1}{2} \right) \right) \\
&\quad - \sum_{t=1}^{T} \left[\frac{\nu+n}{2} \log(1 + \mathbf{x}_t^\top R^{-1} \mathbf{x}_t) + \frac{\nu+1}{2} \sum_{i=1}^{n} \log \left(1 + \frac{x_{ti}^2}{\nu} \right) \right],
\end{aligned}$$

where again $\mathbf{x}_t = (t_\nu^{-1}(u_{t1}), ..., t_\nu^{-1}(u_{tn}))$, and the u_{ti} are the result of applying the empirical marginal transformation to the data.

The above maximization is quite computationally demanding. The matrix inversion is one thing, but that procedure only has to be performed once, before the maximization, as the scale matrix R is held fixed. But the parameter ν appears in the inversion used to obtain the vectors $\mathbf{x}_t$, so every evaluation of the likelihood requires inverting every point in the data set used. In our backtesting example with $n = 8$ and $T = 250$, that means $8 \times 250 = 2000$ inversions of the Student's t distribution in every evaluation of the likelihood. The computational time would thus depend heavily on the time necessary to invert the Student's t distribution. The implementation here uses the `stdtri` method from the `goose`

[1] statistical library. Also, the methods `ludcmp` and `lubksb` from Press et al. (1988) [53] are used to find the determinant of and invert the scale matrix R.

A.3.2 The likelihood for δ for the Clayton copula

Estimating the single parameter δ for the Clayton copula is quite straightforward, as its density has the simple closed-form expression (4.16). The maximization problem is given by

$$\hat{\delta} = \arg\max_{\delta>0} \left[\sum_{t=1}^{T} \log c(\mathbf{u}_t; \delta) \right]$$

where the likelihood $\sum_{t=1}^{T} \log c(\mathbf{u}_t; \delta)$ can be expressed as

$$\begin{aligned} &\sum_{t=1}^{T} \log c(\mathbf{u}_t; \delta) = \\ &\quad T \cdot \left(n \log(\delta) + \log \Gamma\left(\frac{1}{\delta} + n\right) - \log \Gamma\left(\frac{1}{\delta}\right) \right) \\ &\quad - \sum_{t=1}^{T} \left[(\delta + 1) \sum_{i=1}^{n} \log(u_i) + \left(\frac{1}{\delta} + 1\right) \log\left(\sum_{i=1}^{n} u_i^{-\delta} - n - 1 \right) \right]. \end{aligned}$$

The likelihood is thus simple to implement, and its evaluation is relatively cheap, computationally speaking.

A.3.3 The likelihood for δ for the Frank copula

Estimating the parameter δ for the Frank copula is also a case of a simple one-dimensional maximization of a likelihood, but the density is somewhat more complicated to express. As above, we want to find

$$\hat{\delta} = \arg\max_{\delta>0} \left[\sum_{t=1}^{T} \log c(\mathbf{u}_t; \delta) \right], \tag{A.17}$$

where a positive value of δ enables the use of a simulation algorithm of the type outlined in 4.5.2. The density of the multivariate Frank copula is easiest to express through (4.13) and the following expressions, as given in Savu & Trede (2004) [60]:

[1] GNU object-oriented statistics environment: www.gnu.org/software/goose

$$
\begin{aligned}
\phi(t) &= \log\left(\frac{\exp(-\delta t)-1}{\exp(-\delta)-1}\right) \\
\phi'(t) &= \frac{\delta\exp(-\delta t)}{\exp(-\delta t)-1} \\
\phi^{-1(n)}(x) &= \frac{(\exp(\delta)-1)\exp(\delta+x)}{(\exp(\delta)-\exp(\delta+x)-1)^n\delta} \\
&\quad \times \sum_{k=1}^{n-1} B_{n-1,k}\exp((n-1-k)(x+\delta))(\exp(\delta)-1)^{k-1}
\end{aligned}
$$

where

$$B_{i,j} = (i-j-1)B_{i-1,j-1} + jB_{i-1,j},\ B_{i1} = B_{ii} = 1.$$

Algorithm 2 shows the pseudocode for evaluating the density of the Frank copula according to (4.13) and the expressions above.

Algorithm 2 Evaluate Frank copula density

Require: Data $u_1, ..., u_n$, parameter δ

$B \leftarrow n \times n$ empty matrix
for $i = 1, ..., n-1$ **do**
  $B_{i,1} \leftarrow 1\ B_{i,i} \leftarrow 1$
end for
for $i = 3, ..., n-1$ **do**
  for $j = 2, ..., i-1$ **do**
    $B_{i,j} = (i-j-1)B_{i-1,j-1} + jB_{i-1,j}$
  end for
end for
$x \leftarrow \sum_{i=1}^{n} \phi(u_i)$
$y \leftarrow \prod_{i=1}^{n} \phi'(u_i)$
return $\phi^{-1(n)}(x) \cdot y$

The likelihood is then evaluated directly by (A.17) with algorithm 2 called for $t = 1, ..., T$.

A.3.4 Maximization of the likelihoods

For all three likelihoods above, the implementation here uses the bracketing method `mnbrak` and Brent's method `brent` from Press et al (1988) [53] to bracket and minimize the negative likelihood. Brent's method combines a *section search* with *inverse parabolic interpolation*, and does not require derivatives. The method appears robust as far as the calculations here are concerned. To reduce the number of iterations, a routine that uses derivative information might be considered, though this has not been tested here. As for the constraints on the parameter values, for the Frank and Clayton copulae, these were enforced by iterating over values $\tilde{\delta}$ and setting $\delta = \tilde{\delta}^2$. For the Student's t copula, the somewhat unelegant method of setting the negative likelihood to return a high value in the case of a constraint violation was employed. While perhaps not the optimal method, it appeared to work satisfactorily here.

A.4 Calibration and simulation for the Student's t c-vine

The procedure of calibrating and simulating from the Student's t c-vine is somewhat complex, this section will go through some of the details [2].

For the procedure outlined below, we will need the univariate and bivariate Student's t distribution *conditional* on some set of variables. For convenience, this conditional distribution is shown first here. Consider a vector of Student's t distributed variables $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^\top$, where $\mathbf{X} \in \Re^n,\ \mathbf{X}_1 \in \Re^i,\ \mathbf{X}_2 \in \Re^{n-i}$. The distribution of $\mathbf{X}$ is

$$f(\mathbf{x}) = \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)(\nu\pi)^{n/2}|R|^{1/2}} \left(1 + \frac{1}{\nu}(\mathbf{x}-\mu)^\top R^{-1}(\mathbf{x}-\mu)\right)^{-(\nu+n)/2}. \quad \text{(A.18)}$$

Now, we want an expression for the distribution of $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$. First, divide the scale matrix and mean of $\mathbf{X}$ into components corresponding to $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$R = \begin{bmatrix} R_{11}^{i\times i} & R_{12}^{i\times(n-i)} \\ R_{21}^{(n-i)\times i} & R_{22}^{(n-i)\times(n-i)} \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1^{i\times 1} \\ \mu_2^{(n-i)\times 1} \end{bmatrix}.$$

Set

$$\begin{aligned} \mu_{2|1} &= \mu_2 + S_{12}^\top S_{11}^{-1}(\mathbf{x}_1 - \mu_1) \\ R_{2|1} &= \frac{\nu + (\mathbf{x}_1-\mu_1)^\top S_{11}^{-1}(\mathbf{x}_1-\mu_1)}{\nu+i}(S_{22} - S_{12}^\top S_{11}^{-1} S_{12}) \\ \nu_{2|1} &= \nu + i \end{aligned}$$

and insert into the density expression (A.18) to obtain the density of $\mathbf{X}_2|\mathbf{X}_1$. For univariate conditional distributions, $R_{\cdot|\ldots}$ becomes a scalar parameter. The bivariate case is needed for the estimation of the conditional copulae $C_{ij|D_e}$, in which case the matrix $R_{ij|D_e}$ should be scaled by the square roots of its diagonal entries.

Turning to the construction of the vine, the first step is to estimate the necessary unconditional parameters. It turns out that for the n-dimensional c-vine, we need to estimate for each $i \in 2, ..., n$ copula parameters for the sets of variables $\{1, i\}, ..., \{1, ..., i-1, i\}$. Denoting the parameter estimates for a set of variables $\{1, ..., j, i\}$ by $\Theta_{1\cdots ji}$, we thus need to estimate the unconditional parameters as listed in table A.2.

Since the copula used for the vine here is the Student's t copula, each set of copula parameters will be given by $\Theta_{1\cdots ji} = \{R_{1\cdots ji}, \nu_{1\cdots ji}\}$ where as usual $R_{\ldots}$ is the scale matrix and $\nu_{\ldots}$ denotes the degrees of freedom.

[2] The calibration and simulation methodology used is an adaption of the code in Berg & Bakken's `Copulae R` library.

$$\begin{array}{cccc} \Theta_{12} & \Theta_{13} & \cdots & \Theta_{1n} \\ & \Theta_{123} & \cdots & \Theta_{12n} \\ & & \ddots & \vdots \\ & & & \Theta_{12\cdots(n-1)n} \end{array}$$

Table A.2: Unconditional parameter estimates for the c-vine

Now, we want variables $v_1, ..., v_n$ that are uniform on $(0,1)^n$ and dependent according to the vine specification. Start by drawing independent uniforms $u_1, ..., u_n$ on $(0,1)^n$. These are the realized values of the marginal distributions $F_1, ..., F_n$. We have already estimated the bivariate copulae $C_{12}, ..., C_{1n}$, but not the conditional bivariate copulae $C_{ij|D_e}$. From the estimated copula parameters $\Theta_{...}$, we obtain the $C_{ij|D_e}$ by conditioning on the estimated vine distributed variables $v_i,\ i \in D_e$. Finally, from the bivariate copulae C_{ij} and the conditional bivariate copulae $C_{ij|D_e}$, we obtain the univariate conditional distributions $F_{i|...}$ by conditioning on realized values $u_1, ..., u_n$ from the marginal distributions $F_1, ..., F_n$.

Recall the figure showing the c-vine on four variables from chapter 5, shown here again in figure A.1:

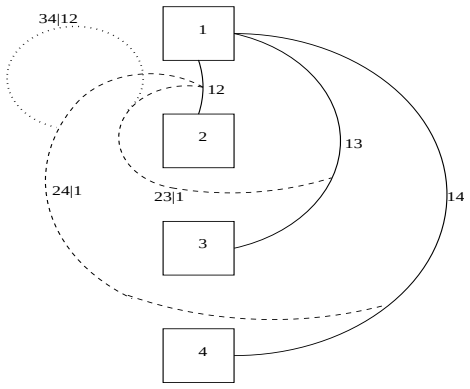


Figure A.1: The c-vine on four variables

The following procedure generates variates v_i with dependence described by the Student's t c-vine.

Calculate v_1:
This is easy:

1. Set $v_1 = u_1$.

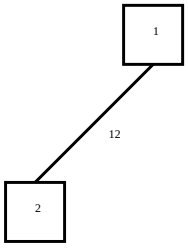


Figure A.2: Calculating v_2

Calculate v_2:
Still quite simple. v_1 and v_2 have a bivariate dependence according to the copula C_{12}, with parameters $\Theta_{12} = \{\nu_{12},\ R_{12}\}$, see figure A.2. Hence,

1. From Θ_{12}, condition on u_1 to calculate $\Theta_{2|1}$, the parameters of the conditional t-distribution $t_{2|1} = t(x_2|x_1 = t^{-1}_{\nu_{12}}(u_1))$. This gives the conditional distribution $F_{2|1}(x) = t_{2|1}(t^{-1}_{\nu_{12}}(x))$ for the vine-distributed variable x.

2. Set $v_2 = F^{-1}_{2|1}(u_2) = t_{\nu_{12}}(t_{2|1}(u_2))$.

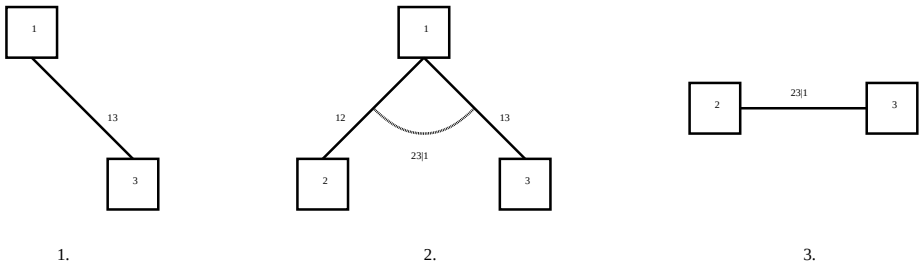


Figure A.3: Calculating v_3

Calculate v_3:
For v_3, we see where the vine strucure enters. v_3 is dependent on v_1 through the copula C_{13}, and is also dependent on v_2 conditional on v_1, through the conditional copula $C_{23|1}$. Figure A.3 illustrates the steps below.

1. From Θ_{13}, condition on u_1 to calculate $\Theta_{3|1}$, the parameters of the conditional t-distribution $t_{3|1} = t(x_3|x_1 = t^{-1}_{\nu_{13}}(u_1))$.
   Let $F_{3|1}(x) = t_{3|1}(t^{-1}_{\nu_{13}}(x))$.

2. From Θ_{123}, condition on v_1 to calculate $\Theta_{23|1}$, the parameters of the conditional t-distribution $t_{23|1} = t(x_2, x_3|x_1 = t^{-1}_{\nu_{123}}(v_1))$. These are the parameters of the conditional bivariate copula $C_{23|1}$. Note that (although v_1 and u_1 are the same in this case) the dependence between variables 2 and 3 is defined as conditional on the vine-distributed variable v_1.

3. From $\Theta_{23|1}$, condition on u_2 to calculate $\Theta_{3|12}$, the parameters of the conditional t-distribution $t_{3|12} = t(x_3|x_1 = t^{-1}_{\nu_{123}}(v_1), x_2 = t^{-1}_{\nu_{23|1}}(u_2))$.
   Let $F_{3|12}(x) = t_{3|12}(t^{-1}_{\nu_{23|1}}(x))$.

4. Set

$$\begin{aligned} v_3 &= F^{-1}_{3|1}(F^{-1}_{3|12}(u_3)) \\ &= t_{\nu_{13}}(t^{-1}_{3|1}(t_{\nu_{23|1}}(t_{3|12}(u_3)))) \end{aligned}$$

We will include the steps for one more variable.

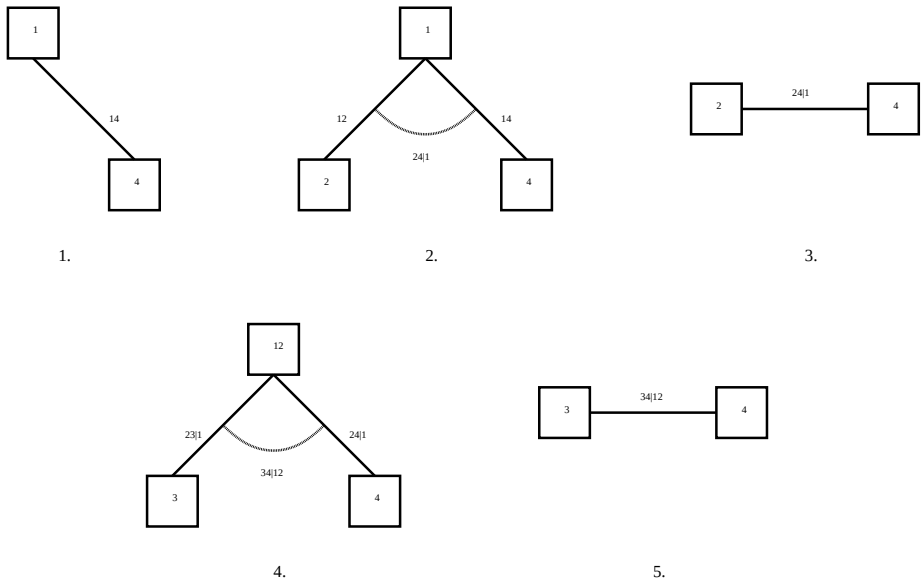


Figure A.4: Calculating v_4

Calculate v_4:
The calculations start to get a bit involved now. See figure A.4 for an illustration of the steps below.

1. From Θ_{14}, condition on u_1 to calculate $\Theta_{4|1}$, the parameters of the conditional t-distribution $t_{4|1} = t(x_4|x_1 = t^{-1}_{\nu_{14}}(u_1))$.
   Let $F_{4|1}(x) = t_{4|1}(t^{-1}_{\nu_{14}}(x))$.

2. From Θ_{124}, condition on v_1 to calculate $\Theta_{24|1}$, the parameters of the conditional t-distribution $t_{24|1} = t(x_2, x_4|x_1 = t^{-1}_{\nu_{124}}(v_1))$. These are the parameters of the conditional bivariate copula $C_{24|1}$.

3. From $\Theta_{24|1}$, condition on u_2 to calculate $\Theta_{4|12}$, the parameters of the conditional t-distribution $t_{4|12} = t(x_4|x_1 = t^{-1}_{\nu_{124}}(v_1), x_2 = t^{-1}_{\nu_{24|1}}(u_2))$.
   Let $F_{4|12}(x) = t_{4|12}(t^{-1}_{\nu_{24|1}}(x))$.

4. From Θ_{1234}, condition on v_1, v_2 to calculate $\Theta_{34|12}$, the parameters of the conditional t-distribution $t_{34|12} = t(x_3, x_4|x_1 = t^{-1}_{\nu_{1234}}(v_1), x_2 = t^{-1}_{\nu_{1234}}(v_2))$. These are the parameters of the conditional bivariate copula $C_{34|12}$.

5. From $\Theta_{34|12}$, condition on u_3 to calculate $\Theta_{4|123}$, the parameters of the conditional t-distribution $t_{4|123} = t(x_4|x_1 = t^{-1}_{\nu_{1234}}(v_1), x_2 = t^{-1}_{\nu_{1234}}(v_2), x_3 = t^{-1}_{\nu_{34|12}}(u_3))$
   Let $F_{4|123}(x) = t_{4|123}(t^{-1}_{\nu_{34|12}}(x))$.

6. Set

$$\begin{aligned} v_4 &= F^{-1}_{4|1}(F^{-1}_{4|12}(F^{-1}_{4|123}(x))) \\ &= t_{\nu_{14}}(t^{-1}_{4|1}(t_{\nu_{24|1}}(t^{-1}_{4|12}(t_{\nu_{34|12}}(t^{-1}_{4|123}(u_4)))))). \end{aligned} \tag{A.19}$$

The procedure generalizes to higher dimensions, for this illustration the first four variables will suffice.