Modelling dependence of interest rates, inflation rates and stock market returns

Hans Waszink AAG MBA MSc
Waszink Actuarial Advisory Ltd.
Sunnyside, Lowden Hill, Chippenham, UK
hans@hanswaszink.com

Abstract:

In the first part of this article, an approach to model the value of an outstanding, discounted liability under the impact of uncertain interest and inflation rates is discussed. Interest and inflation rates are modeled separately as time series to take into account autocorrelation. Subsequently, the dependence between interest and inflation is modeled using copulas. The goodness of fit of some copulas can be evaluated on the basis of historic data using a quantile plot. This is done for the Gumbel, Clayton and Independent copulas. The Gumbel copula, which gives the best fit, is then compared with the Normal copula to show that the two copulas are very similar with the parameters chosen. The distribution of the required reserve is shown under four different copula assumptions: comonotonicity, which represent the best case, countermonotonicity which represents the worst case, and the Gumbel and Normal copulas which represent more realistic scenarios.

The choice of copula has considerable impact on the higher percentiles of the required reserve, and the adopted approach is effective in selecting a suitable copula for the modelling of two underlying variables.

In the second part of the article, the application of copulas to the modelling of three dependent variables (interest, inflation and stock market return) is investigated. As the Clayton and Gumbel copulas only have a single parameter, they prove to be less suitable for the modelling of more than two dependent variables. The normal copula appears more suitable in this case as the dependence between each modelled variable can be set separately by a different parameter in the correlation matrix.

Keywords: copulas, dependence, autocorrelation, time series, stochastic modelling, goodness of fit, interest, inflation, stock market returns, discounting.

1. INTRODUCTION

An insurance company has liabilities on its books for the payment of future insurance claims. The present value of the liabilities is unknown for two main reasons: future inflation is unknown, and future interest rates are unknown. In order to quantify these uncertainties, the probability distribution of the present value of the future liabilities is estimated.

The estimation of the probability distribution requires to make assumptions regarding the following:

- The distribution of interest and inflation rates in future periods.
- The dependence between interest (inflation) rates in one future period with the interest (inflation) rate in another period. This type of dependence is generally referred to as autocorrelation.
- The dependence between inflation and interest rates in the same period. Although linear correlations are often used to model dependencies, their pitfalls are well-known[1]. Therefore we will investigate the use of other types of dependence structures, known as copulas.

The aim of this study is to investigate the usefulness and limitations of different types of dependence structures (so called copulas) for the estimation of the distribution of the present value of the liabilities. Particular focus is placed on the materiality of various assumptions regarding the dependence structure, as well as practical complexities in the use of the modelling techniques applied.

The study comprises some introductory theory, followed by a case study. In the first part of the case study we consider the combined effects of interest and inflation rates on the present value of a liability. In the second part of the case study, we will assume that part of the assets backing the liabilities are invested in common stock. This requires the combined modelling of interest and inflation rates and stock market returns.

Section 2 contains a brief description of a simple time series model and some of its properties. In section 3, the present value of the liabilities is defined in terms of the underlying variables, and the uncertainty contained therein. Section 4 then introduces time series models for interest and inflation rates, while section 5 shows an overview of the data used. Section 6 discusses the parameterisation of the time series models. Section 7 describes a method to model the dependence between interest and inflation rates using copulas. In section 8, the technique described in section 7 is applied to the problem at

[1] See Embrechts et al.(1999)

hand. Section 9 contains conclusions of the results obtained in the first part of the case study. In section 10, the model is extended by introducing stock market returns as another variable. Section 11 finally contains overall conclusions.

The author would like to thank the Correlation Working Party of the 2006 CAS Winter Forum[2], as well as Catalina Stefanescu of London Business School, for useful comments provided.

2. TIME SERIES

Time series models are commonly used for variables of which observations are available sequentially in time, and consecutive observations are dependent. Both these properties typically apply to interest as well as inflation rates.

A simple example of a time series is an autoregressive process of order 1 (denoted by AR(1)), which is given below:

$$X(t) = a + bX(t\text{-}1) + \varepsilon(t), \quad t = 1,\ldots,T \qquad (1)$$

with

$X(.)$: array of stochastic variables, $t = 0,1,\ldots,T$, $X(0)$ a given constant.
$\varepsilon(t)$: independent random error within period $(t\text{-}1,t)$, with $N(0,\sigma)$ distribution.
a,b: model parameters.

A time series is called *stationary* when the distribution of $X(t)$ and dependence structure of X do not change over time[3], hence is independent of t. This means (amongst other things) that $\mathrm{Var}(X(t)) = \mathrm{Var}(X(\tau))$ for each t, $\tau = 1,2,\ldots,T$.

For a stationary AR(1) time-series, satisfying equation *(1)* above, we have that:

$$\mathrm{Var}(X(t)) =$$

$$\mathrm{Var}(a + bX(t\text{-}1) + \varepsilon(t)) =$$

$$\mathrm{Var}(bX(t\text{-}1)) + \mathrm{Var}(\varepsilon(t)).$$

[2] See http://www.casact.org/pubs/forum/06wforum/
[3] See Box et al.(1994)

In this way the variance of $X(t)$ can be broken down into two components: the variance explained by the previous observation $X(t-1)$, and the residual variance Var ($\varepsilon(t)$)

From the equality Var($X(t)$) = Var ($X(t-1)$) it follows that:

$$\mathrm{Var}(X(t)) =$$

$$\mathrm{Var}\ (bX(t-1)) + \mathrm{Var}\ (\varepsilon(t))=$$

$$b^2\ \mathrm{Var}(X(t)) + \mathrm{Var}\ (\varepsilon(t)).$$

so that:

$$\mathrm{Var}(X(t))(1- b^2) = \mathrm{Var}\ (\varepsilon(t)).$$

Hence the residual variance Var ($\varepsilon(t)$) is a fraction ($1- b^2$) of the total variance Var($X(t)$).

The correlation between $X(t)$ and $X(t-1)$ is derived as follows
(using Var($X(t)$) = Var ($X(t-1)$)).

$$\rho(X(\mathrm{t}),X(\mathrm{t}-1)) =$$

$$(\mathrm{Cov}\ (X(\mathrm{t}),X(\mathrm{t}-1))/\mathrm{Var}\ (X(t)) =$$

$$\mathrm{Cov}\ (bX(t-1) + \varepsilon(t), X(t-1))/\mathrm{Var}\ (X(t)) =$$

(as $\varepsilon(t)$ and $X(t-1)$ are independent)

$$\mathrm{Cov}\ (bX(t-1), X(t-1))/\mathrm{Var}\ (X(t)) =$$

$$b\ \mathrm{Var}(X(t-1))\ /\mathrm{Var}\ (X(t)) =$$

$$b\ .$$

Similarly it can be shown that the correlation between $X(t)$ and $X(t-2)$ is b^2, or more general:

$\rho(X(t),(X(\tau)) = b^{|t-\tau|}$ for each $t,\tau = 1,2,\ldots,T$.

The parameter b can be estimated on the basis of historic observations of the variable $X(t)$. If the estimate of b is not significantly different from 0, then there is no evidence of autocorrelation and the $X(t)$ are assumed to be uncorrelated.

3. OUTSTANDING LIABILITY UNDER UNCERTAIN INTEREST AND INFLATION RATES

We consider the value of an outstanding claims reserve as the present value of inflated and discounted future claim payments. Interest and inflation rates are modeled as random variables. As a starting point, we use uninflated projections of future claim payments in each future payment period. These can be derived from actuarial reserving methods which contain a projection of future claim payments before taking into account the impact of claims inflation.

It is assumed that there is a well defined asset portfolio that serves to pay the future liabilities. This can be a portfolio of government bonds, or a mix of bonds and common stock, or another asset portfolio. In the first part of the case study, we will assume that the asset portfolio consists of government bonds generating the three year market interest rate on US government bonds. This is obviously a simplification of reality as we assume that the asset portfolio profile will remain constant over the full run-off period of the liabilities.

We define:

$C(t)$: Uninflated, fixed and given cashflow projection at time t.
$Inf(t)$: Inflation rate in period $(t,t+1)$, $t = 0,1,2,\ldots$
$Int(t)$: Interest rate in period $(t, t+1)$, $t = 0,1,2,\ldots$

$Ac(t)$: Actual cashflow at time t.

$Ac(t)$ is equal to:

$$Ac(t) = C(t) \times \prod_{\tau=0}^{t-1} [1 + Inf(\tau)], \quad t = 1,2,3,\ldots$$

For simplicity it is assumed that $Ac(t)$ is the product of the cashflow projection $C(t)$, which is fixed and given, and future inflation rates only. Therefore the only uncertain factor in actual future cashflows is future inflation which can represent general inflation, or a rate of inflation specific to an economic sector or line of insurance products. In this study we have used medical inflation, a line-specific inflation impacting on health insurance related liabilities.

The inflation rates represent a component of systematic risk in the cash flow projection, i.e. they affect all individual claims simultaneously and to the same extent. To relax the assumption that inflation is the only uncertain factor affecting future cashflows, additional components of unsystematic risk can be added without any difficulty, however these are excluded here.

We further define

Df(t): Discount factor in period (*t,t+1*), *t=0,1,2,…*:

$$Df(t) = \frac{1}{1 + Int(t)}$$

RR(t): Required reserve at time *t* , *t=0,1,2,…*:

.

$$RR(t) = \sum_{s>t} [Ac(s) \times \prod_{\tau=0}^{s-t-1} Df(t+\tau)]$$

The required reserve is the total of all actual future cashflows discounted at actual future interest rates. Obviously *RR(t)* is not known in advance as it is a function of *C(t)*, *Inf(t)* and *Int(t)* with future interest and inflation rates unknown.

The distribution of the *RR(t)* is a function of the marginal distributions of the interest and inflation rates after time *t* and the dependencies between interest rates in different periods, the dependence between inflation rates in different periods, and the dependence between inflation and interest rates in the same period and in different periods.

In order to derive the distribution of *RR(t,* and in particular *RR(0)*, we will first estimate the distributions of interest rates and inflation rates respectively, and then pay attention to the dependence between them.

4. MODELLING THE DISTRIBUTION OF INTEREST AND INFLATION RATES

In the two paragraphs of this chapter, the estimation of the distributions of interest and inflation rates are discussed.

4.1 Interest rates

A discrete version of the CIR[4]-model for a single interest rate is used. Although not the only interest rate model nor the most sophisticated one, finding the best possible interest rate model is outside the scope of this report. We will therefore simply assume that the CIR–model is adequate for the purpose although we will check the model fit.

The discrete CIR-model is a time-series model of the following form:

$$Int(t) = max\{0, Int(t-1) + a[b - Int(t-1)] + \sqrt{Int(t-1)}\varepsilon_{int}(t)\} \qquad (2)$$

with

$Int(t)$:	the interest rate in the period $(t,t+1)$
a:	the average speed of reversion to the long term mean interest rate;
b:	the long term mean interest rate.
$\varepsilon_{int}(t)$:	random deviation in period $(t,t+1)$. The $\varepsilon_{int}(t)$ are mutually independent with marginal distributions $N(0,\sigma^2)$.

The '*max*' formula is used in order to prevent that the distribution of *Int(t)* includes negative values, although this normally occurs with only a very small probability. The CIR-model is well known and has several desirable properties such as:

- Interest rates are mean reverting;
- Interest rates are non-negative.
- Interest rates are heteroskedastic, i.e. variance increases with mean.
- Interest rates at adjacent points in time are correlated.
- Confidence intervals widen for interest rates projections further into the future.

[4] Cox Ingersoll Ross, see Kaufmann et al.(2001)

4.2 Inflation rates

For inflation rates we will investigate the suitability of a second order autoregressive process (denoted by AR(2)) of the following form:

$$Inf(t) = c_0 + c_1\, Inf(t-1) + c_2\, Inf(t-2) + \varepsilon_{inf}(t) \qquad (3)$$

with

c_0 , c_1 , c_2 : model parameters.
$\varepsilon_{inf}(t)$: random deviations in period $(t,t+1)$.

The $\varepsilon_{inf}(t)$ are mutually independent with identical marginal distributions $N(0,\sigma)$.

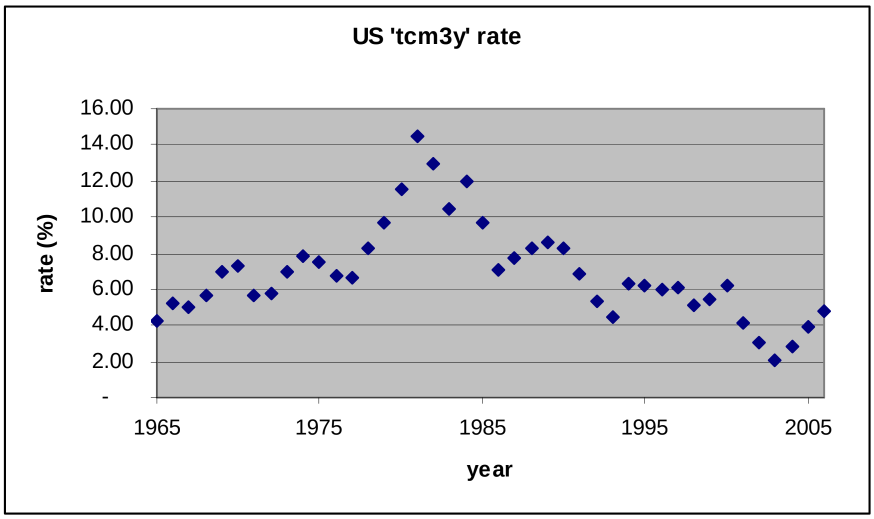
Properties of the AR(2) model that make it potentially suitable for the modelling of inflation rates are:

- If $c_2 < 0$, rates may exhibit cyclicality.
- Observations at adjacent points in time are correlated.
- Confidence intervals widen for projections further into the future.
- If the estimated parameter c_2 is not statistically significant, then the AR(2) model reduces to an AR(1) model with $c_2 = 0$.

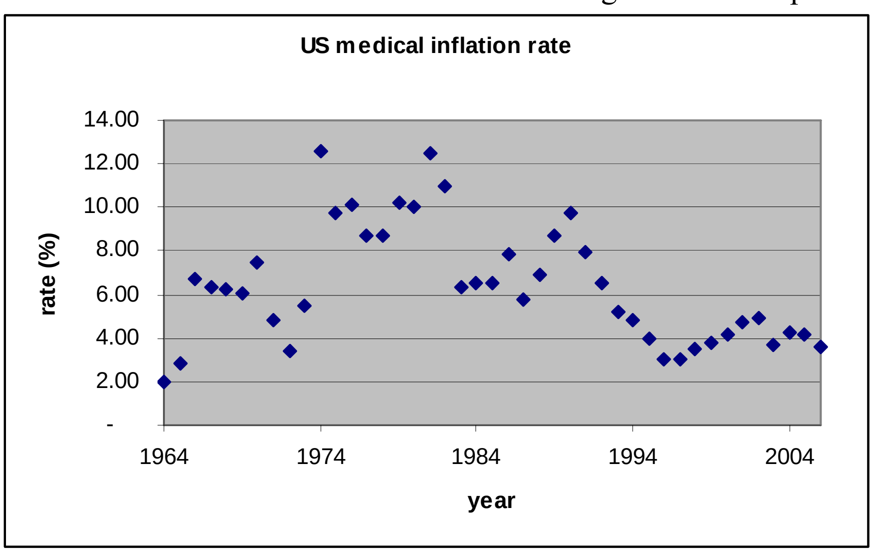
5. DATA

For our analysis, we have used the following data of interest rates, medical inflation rates and stock market returns. Medical inflation represents the increase of overall costs of medical care, and is of interest to health insurance companies and medical coverage providers. Data sources can be found in appendix I.

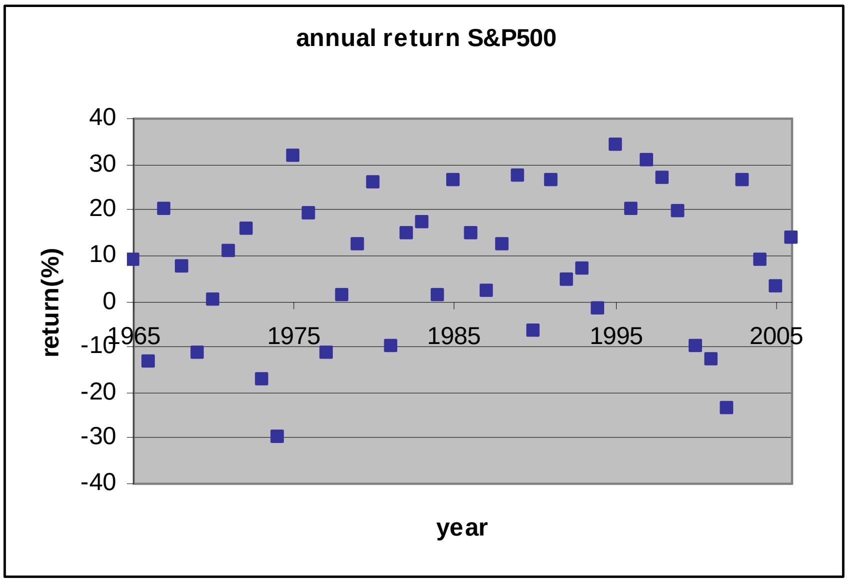
- Annual 3 year interest rates on US government securities over the period 1965 to 2006.



- Increase in annual US medical care index figures over the period 1964 to 2006.



- Total annual return of the S&P 500 index including reinvestment of dividends over the period 1966 to 2006.



6. PARAMETERISATION OF INTEREST AND INFLATION MODELS

6.1 Parameterisation of the CIR model

The CIR model can be regarded as a variation of the AR(1) model. The expected interest rate in period *t* is an almost linear function (linear as long as interest rates remain positive) of the interest rate in the previous time period. The error term in *Int(t)* is equal to $\sqrt{Int(t-1)}\varepsilon_{int}(t)$, thus allowing for an element of heteroskedasticity, i.e. the variance increases with the interest rate in the previous period.

The parameters of the CIR model can be estimated by minimizing the sum of squares of the residuals over the observed time period. In other words, *a* and *b* are such that the following expression is minimised over the observed period:

$$\sum \left\{\frac{(Int(t) - Int(t-1) - a[b - Int(t-1)]}{\sqrt{Int(t-1)}}\right\}^2 =$$

$\sum \varepsilon_{\text{int}}(t)^2$ if $Int(t-1) + a[b - Int(t-1)] + \sqrt{Int(t-1)}\varepsilon_{int}(t)$ is positive for each observed *t*.

The latter condition needs to be verified once the solutions for *a* and *b* are found.

We are aware that this estimation procedure may lead to estimates that are inconsistent with current market interest rates and yield curve. The CIR model can be parameterised on the basis of the current yield curve alone[5], but by doing so we would ignore all historic data altogether. However, the purpose of this case study is to investigate the dependence structure of interest rates as well as the dependence between interest rates, inflation and stock market returns. This requires the assumption that historic interest rates and their dependence can be used as an estimate for the distribution of future rates, and so we will use historic rates as the basis for the parameterisation.

We have found the following values for *a*, *b* and *σ* using the Solver function in Excel:

a	0.086
b	0.067
σ	0.005

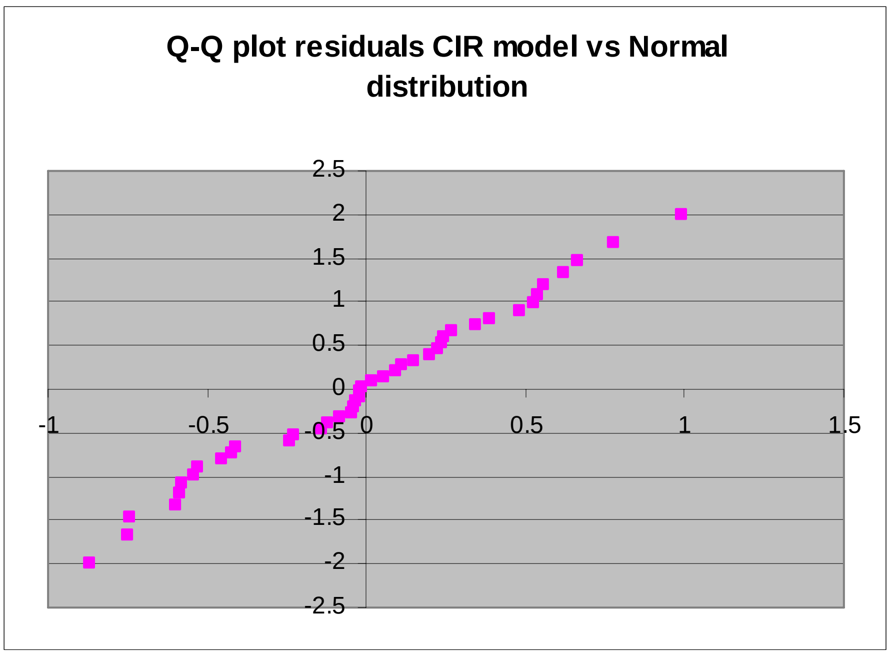
[5] See Kaufmann et al.(2001)

The R^2 of the regression is 55%.

R^2 is the percentage of variance explained by the regression,

$R^2 = 1\text{-}Var(\varepsilon_{int}(t))/Var(Int(t))$.

The quantile plot of the residuals against the normal distribution looks as follows:



The plot shows a pattern fairly similar to a straight line which indicates that the normal distribution gives a reasonably good fit.

We have also performed three goodness-of fit tests to determine whether the normal distribution is suitable for the residuals. These test are the Chi-squared test, the Kolmogorov–Smirnov, and the Anderson-Darling test. The Chi-squared test is a procedure to verify the goodness of fit across the entire range of the distribution. A drawback of the test is the fact that the range of outcomes is divided up into bins, where the choice for the number of bins and their boundaries contains a degree of arbitrariness. The Kolmogorov–Smirnov test does not require binning which makes it less arbitrary then the Chi-squared test, but a drawback of this test is that it does not detect tail discrepancies very well. The Anderson-Darling test finally does not require binning either, but focuses on the tails of the distribution while the Kolmogorov–Smirnov test focuses on the middle of it.

We have performed the three tests with 'Bestfit' software[6]. The p-values of the tests performed on the sample of residuals are as follows:

Table 1: Goodness of fit test results (number of points: 41)

Test	**p-value**
Chi-squared	0.8939
Kolmogorov-Smirnov	>0.25
Anderson-Darling	>0.15

We can conclude that none of the three tests rejects the hypothesis of normality of residuals at the 5% or even the 10% level.

6.2 Parameterisation of the AR(2) model for inflation rates

A procedure for the estimation of the parameters of the AR(2) model can be found in Box et al.(1994).

First compute the first and second order autocorrelation coefficients as follows:

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$, $k = 1,2$ and n is the number of observations x.

The parameters in equation *(3)* are estimated by:

$$c_1 = \frac{r_1(1-r_2)}{1-r_1^2} ,$$

$$c_2 = \frac{r_2 - r_1^2}{1-r_1^2} .$$

[6] See the 'Bestfit' manual for an exact definition of the three tests. The Chi-squared test has been performed using 8 equiprobable bins.

If the parameter c_2 is very close to, and not significantly different from 0, there is no cyclical pattern in the correlation structure and the process is virtually identical to an AR(1) process. In that case, c_2 =o which implies $r_2 = r_1^2$ which is exactly the correlation structure for an AR(1) process described in section 2.

Box et al.(1994) also give the following (asymptotic) estimations of the standard errors of the estimates for c_1 and c_2 :

$$\text{s.e. } (c_1) = \text{ s.e. } (c_2) = \sqrt{\frac{1-c_2^2}{n}} .$$

Using the data of historic medical inflation rates and using the assumption of stationarity of the process, we find the following estimates:

parameter	estimate	standard error
c_0	1.219	
c_1	0.871	0.136
c_2	-0.086	0.136

The value of c_0 is found by assuming that the process is stationary, therefore the historic average is set equal to the expectation of future inflation rates. This assumption requires that:

$$c_0 = \frac{1}{n}\sum_{t=1}^{n} Inf(t)(1-c_1-c_2) .$$

The derivation of the standard error of c_0 is not straight forward, as c_0 is a function of the other parameters as well as the historic mean. We will therefore not attempt to calculate it, but instead set c_0 such that the condition of stationarity is satisfied. Assuming a different value, including 0, of c_0 would imply that the process will converge to a lower expected value so that the condition of stationarity would be violated.

The standard errors of c_1 and c_2 are equal to 0.136. As the parameter c_2 is very small and not significant, we will use an AR(1) process instead of an AR(2) process.

The parameter estimates for the AR(1) process now become:

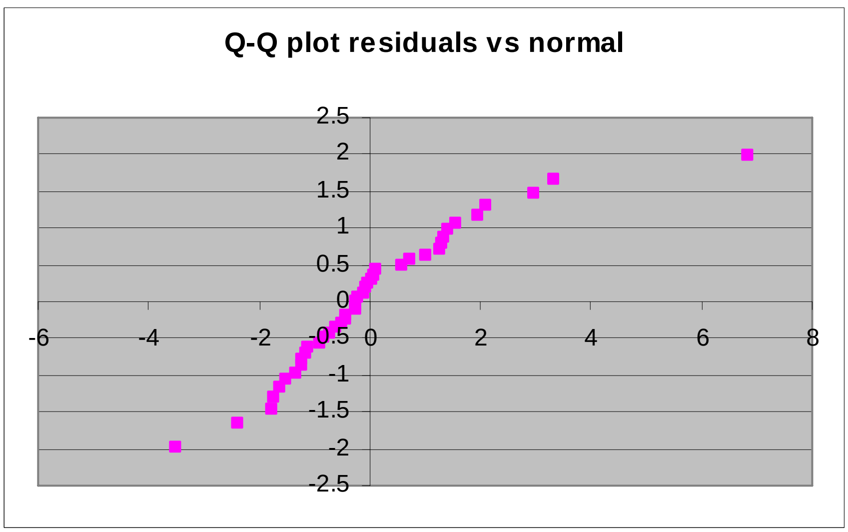
parameter	estimate	Standard error
c_0	1.122	
c_1	0.802	0.079

The standard error of c_1 for an AR(1) process equals (also see Box et al.(1994)):

$$\text{s.e. } (c_1) = \sqrt{\frac{1-c_1^2}{n}}$$

As the parameter estimate is more than ten times the standard error, it can be concluded that the parameter c_1 is significant at 5% level.

The Q-Q plot of the residuals is as follows:



The plot shows a clear deviation from a straight-line pattern in the upper right section of the graph which indicates that the normal distribution potentially does not give a good fit. Again we performed the three goodness-of-fit tests which give the following results:

Table 2: Goodness-of fit test results (number of points: 41)

Test	**p-value**
Chi-squared	0.0621
Kolmogorov-Smirnov	<0.025
Anderson-Darling	<0.01

The latter two tests reject the hypothesis of normality of the distribution at the 5% level, which confirms the impression from the QQ plot that the normal distribution is not a good fit.

In an attempt to improve the fit of the distribution we performed a logarithmic transformation of the observed inflation rates, and again try to fit an AR(2) model. The new model then becomes:

$$ln[Inf(t)] = c_0 + c_1\, ln[Inf(t-1)] + c_2\, ln[Inf(t-2)] + \varepsilon_{l\_inf}(t)$$

Parameter estimates and R^2 are:

Table 3: Parameter estimates logtransfrom AR(2) model

	parameter estimate	standard error
c_0	1.064	
c_1	0.254	0.156
c_2	0.143	0.156
R^2	0.603	

We see that neither c_1 nor c_2 are significant so refit the model again omitting the second-order component c_2 . Results are as follows:

Table 4: Parameter estimates logtransfrom AR(1) model

	parameter estimate	standard error
c_0	0.417	
c_1	0.764	0.158
R^2	0.583	

The QQ-plot against a normal distribution and results of the goodness of fit tests are given below:

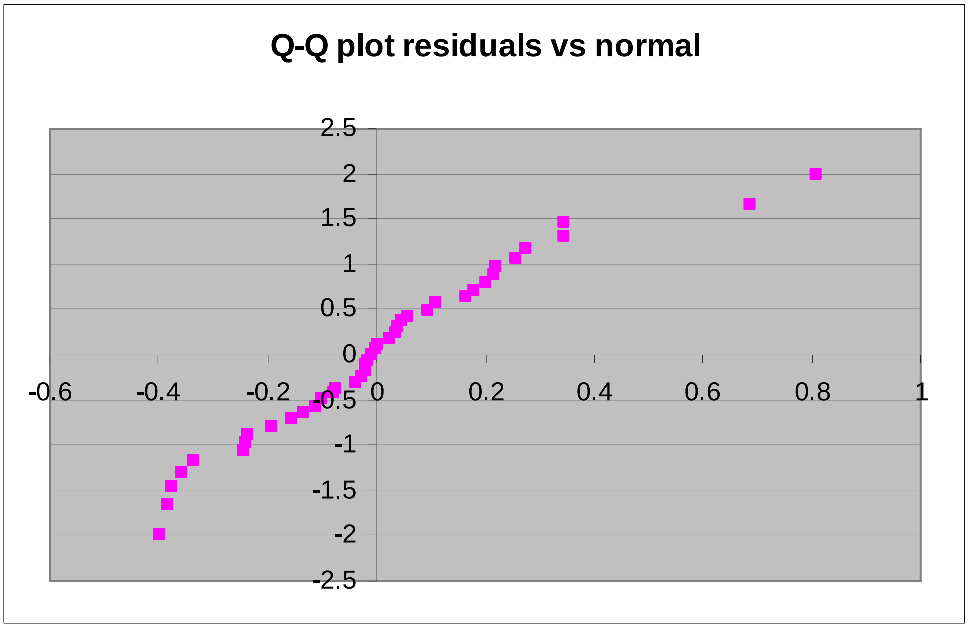


Table 5: Goodness-of fit test results (number of points: 41)

Test	**p-value**
Chi-squared	0.8567
Kolmogorov-Smirnov	>0.1
Anderson-Darling	>0.15

Although the QQ plot shows some deviation from a straight line pattern, none of the three test reject the hypothesis of normality at the 5% or even the 10% level, hence we will assume that the normal distribution is suitable.

7. MODELLING DEPENDENCE USING COPULAS

Given the marginal distribution functions of two or more stochastic variables, a copula fully defines the dependence between the variables. A copula C is the probability distribution function of p uniform random variables U on the unit interval:

$C(u_1,u_2,.., u_p) = P(U_1 \leq u_1 , U_2 \leq u_2 ,..., U_p \leq u_p)$.

The function $C(F_1 (x_1), F_2 (x_2),..., F_p (x_p)) = P(X_1 \leq x_1 , X_2 \leq x_2 ,..., X_p \leq x_p)$

defines a multivariate distribution with marginal distributions $F_1 , F_2, ..., F_p$ and the dependence structure of the variables $X_1 , X_2, ..., X_p$ defined by the copula C.

Although the concept of the copula is theoretically sound, the actual parameterisation of most copula families on the basis of real data as well as verifying the goodness-of-fit remains problematic[7]. Frees and Valdez(1998) suggest a procedure for copula fitting using graphical inspection through a QQ-plot for the family of so-called Archimedean copulas.

Two-dimensional Archimedean copulas are of the form:

> $C_{\varphi}(u,v) = \varphi^{-1}(\varphi(u)+ \varphi(v))$ with $0 < u,v \leq 1$ and φ a convex decreasing function with domain (0,1].

For two random variables X and Y with dependence defined by the Archimedean copula C_φ, it can be shown that the random Variable $Z = C_\varphi(F_X(X), F_Y(Y))$ has the following distribution function:

$$F_Z(z) = z- \varphi(z)/\varphi'(z).$$

This implies that, assuming the dependence between X and Y is described by a given Archimedean copula C_φ , the variable Z should follow the distribution function given above. Hence comparing n ordered (pseudo)-observations of Z with the percentiles of the distribution function of Z in a Q-Q plot allows for inspection of the goodness of fit of the assumed distribution of Z hence of the copula function C_φ. The observations of Z are derived from the observations of X and Y and the relation $Z = C_\varphi(F_X(X), F_Y(Y))$. An estimation procedure for the estimation of the pseudo observations of the values of Z is also outlined in Frees and Valdez(1998).

---

[7] See for example Mikosch (2005) for a discussion of disadvantages of copulas.

The interpretation of the Q-Q plot is no different than the Q-Q plot for any other single random variable. The closer observations are to the corresponding percentiles of the theoretical distribution, the better the fit of the distribution. Hence a Q-Q plot showing a pattern close to the straight line through the origin and (1,1) indicates a good fit of the copula to the data.

We apply this technique to fit the dependence between $\varepsilon_{I\_inf}(t)$ and $\varepsilon_{int}(t)$, comparing three types of Archimedean copulas: the Gumbel, Clayton and Independent copulas as given below.

Gumbel:	$C(u,v) = \exp\{-[(-\ln u)^{\alpha} + (-\ln v)^{\alpha}]^{1/\alpha}\}$,	$\varphi(u) = (-\ln u)^{\alpha}$
Clayton:	$C(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$,	$\varphi(u) = u^{-\alpha} - 1$
Independent:	$C(u,v) = uv$,	$\varphi(u) = -\ln u$.

The 'Independent' copula is simply assuming that the residuals of the interest and inflation rates are independent, and does not require further parameterisation. For the Gumbel and Clayton copulas, we have estimated the parameter α by plotting the QQ-plot in Excel, and minimising the sum of the squared differences between the observations and the fitted values[8].

[8] Parameters can also be found using maximum likelihood estimation, see Frees and Valdez (1998)

The results are as follows:

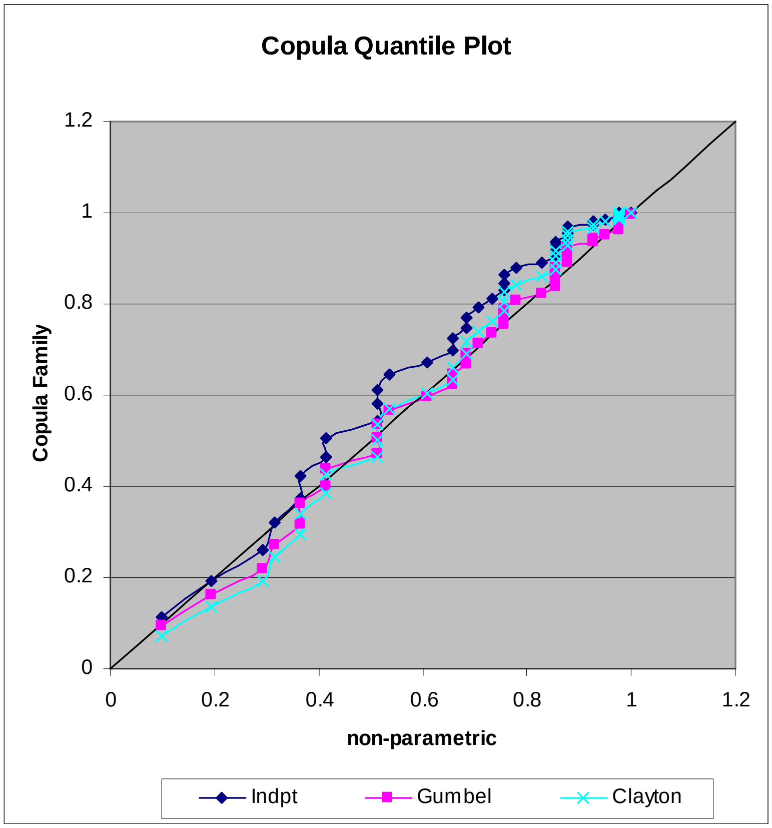


Table 6: Copula parameterisation

Copula	Parameter α	Sum of squared distances
Independent	n/a	0.169048
Gumbel	1.27	0.025189
Clayton	0.38	0.080636

From the plot and the values of the squared distances it becomes apparent that the Gumbel copula gives the best fit of the three copulas used. It can be expected that there is a dependency between $\varepsilon_{int}(t)$ and $\varepsilon_{I\_inf}(t)$ as changes in both inflation and interest rates are driven by the same or related macro-economic factors. This expectation is confirmed by the QQ-plot of the copulas which shows there are better fitting copulas than the Independent copula.

It is interesting to note in this regard that the observed linear correlation between $\varepsilon_{l\_inf}(t)$ and $\varepsilon_{int}(t)$ equals 0.19, with a standard error of 0.15. Hence, although the linear correlation between the two variables is not significantly different from 0, the copula parameterisation procedure clearly suggests that the two variables are not independent.

8. APPLICATION OF THE TIME SERIES AND COPULA MODELS

We have now defined time series models for interest and inflation rates together with the copula representing the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$. As a result the joint distribution of interest and inflation rates is also fully defined. As *RR(t)* is fully determined by the deterministic uninflated cashflows *C(t)* in combination with interest and inflation rates during the projection period, the distribution of all *RR(t)* is fully defined by the joint distribution of inflation and interest rates and *C(t)* . The distribution of *RR(t)* is derived by means of simulation.

For the uninflated cashflow projection *C(t)* we set $C(t) = 1$ for $t= 1,2,\ldots,10$ and 0 otherwise. For the choice of the copula defining the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$, several alternative scenarios are investigated:

1. $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ are comonotonic, i.e. the dependence between the two is maximum. As both $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ are Normal random variables, the linear correlation between them is 100%. This is the best case scenario for the insurer with respect to the dependence between the two error terms. The underlying assumption is that random deviations of interest rates are fully correlated with random deviations of inflation rates, hence unexpected increases in inflation are always accompanied by unexpected increases in interest rates. As increases in inflation rates lead to increases in *RR(.)* whereas increases in interest rates lead to decreases of *RR(.)*, the comonotonic assumption implies that there always is a compensating effect of the two random errors on the liability for the insurer. Therefore this scenario represents a best case for the insurer with respect to the occurrence of extremely high values of *RR(t)*.

2. In the second scenario, the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ is assumed to be 'countermonotonic'[9], meaning unexpected increases in inflation rates are always

[9] Characterization of comonotonicity and countermonotonicity can be found in Denuit and Dhaene (2003)

accompanied by unexpected decreases in interest rates. Contrary to the first alternative, this scenario represents the worst case with respect to the occurrence of extremely high values of *RR(.)*, as the effects of unexpected inflation in any particular period are aggravated by lower interest earnings in the same period.

3. In the third scenario, the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ is parameterised using the Gumbel copula as discussed in the previous section.

4. In the fourth alternative, the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ is modeled as a multivariate Normal distribution, with the dependence between the two random variables fully characterised by their linear correlation coefficient.

The simulated results of each of the four methods are shown in Figure 1 below[10], with BC (Best Case), WC (Worst Case), Gumbel and Normal depicting *RR(0)* in alternatives 1-4 respectively. The right tail is shown in more detail in figure 2.

[10] Results were generated using IGLOO software.

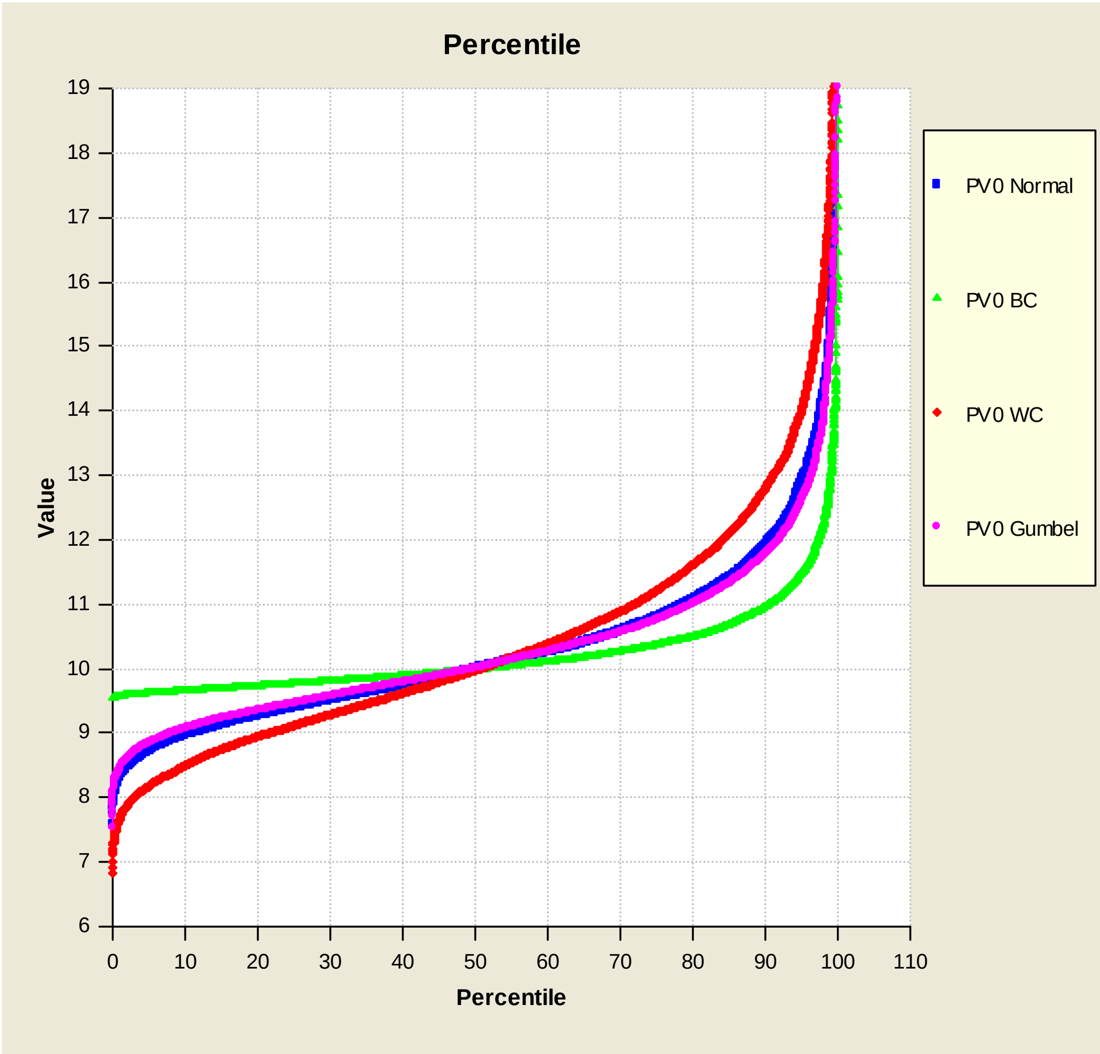


Figure 1: Simulated distributions of RR(0) scenario 1-4

In alternative 4, a linear correlation between $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ of 0.19 is applied, the historically observed correlation between the residuals. The Gumbel copula in alternative 3 gives rise to the same linear correlation.

The right tail of the distributions are shown below.

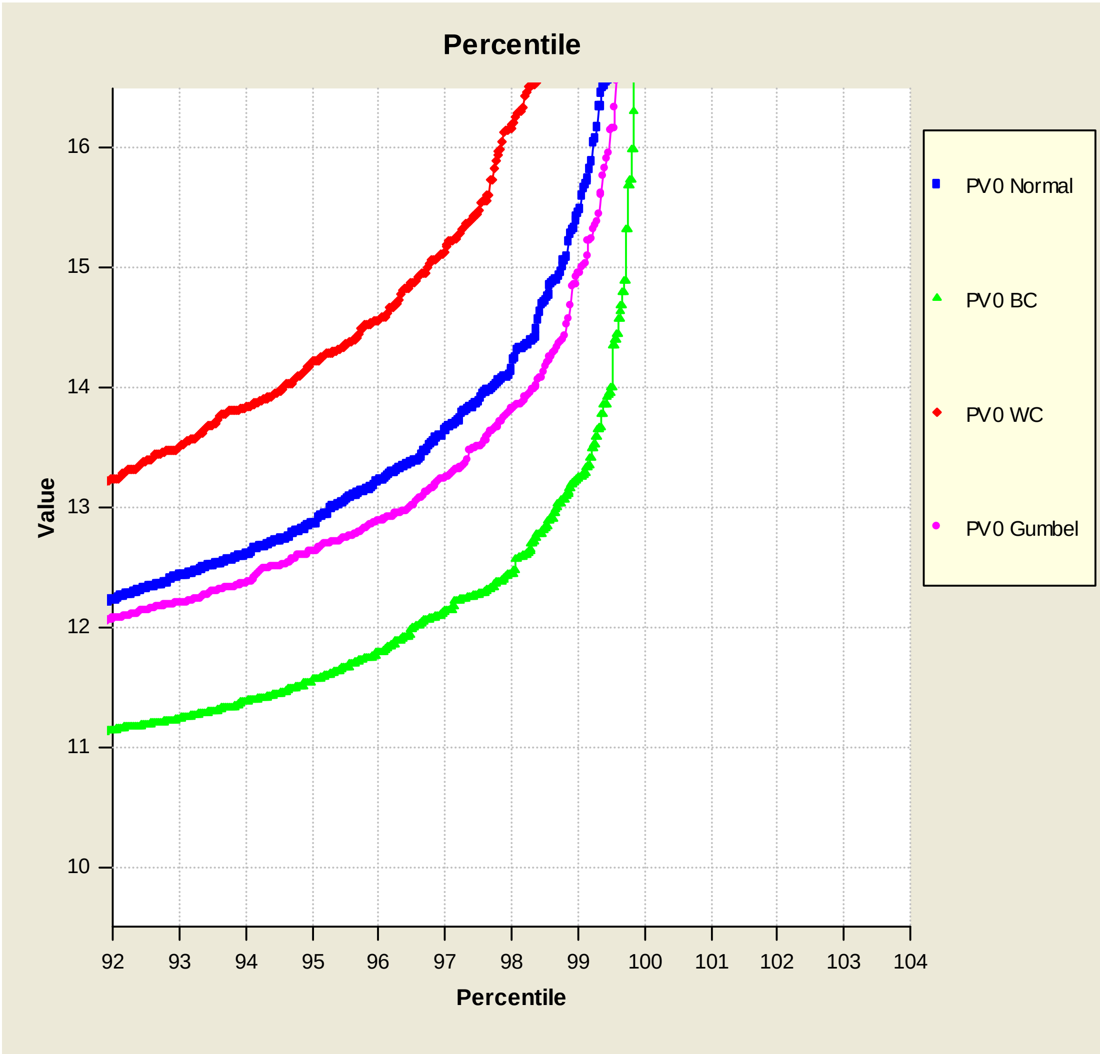


Figure 2: Right tail of simulated distributions of RR(0)

The fact that the difference between the distributions under the two Gumbel and Normal copulas is small suggests that the two copulas generate very similar dependence structures. This is confirmed by the simulated rank scatter plots of the two copulas shown in Figure 3 below.

A rank scatter plot shows simulated pairs of uniform random variables under a given dependence structure between the two variables. When realizations are spread evenly across the square, this indicates a low degree of dependence. A high degree of dependence is indicated by concentrations of points in certain parts of the square. For example tail correlation leads to a higher concentration of realizations in the corners of the square.

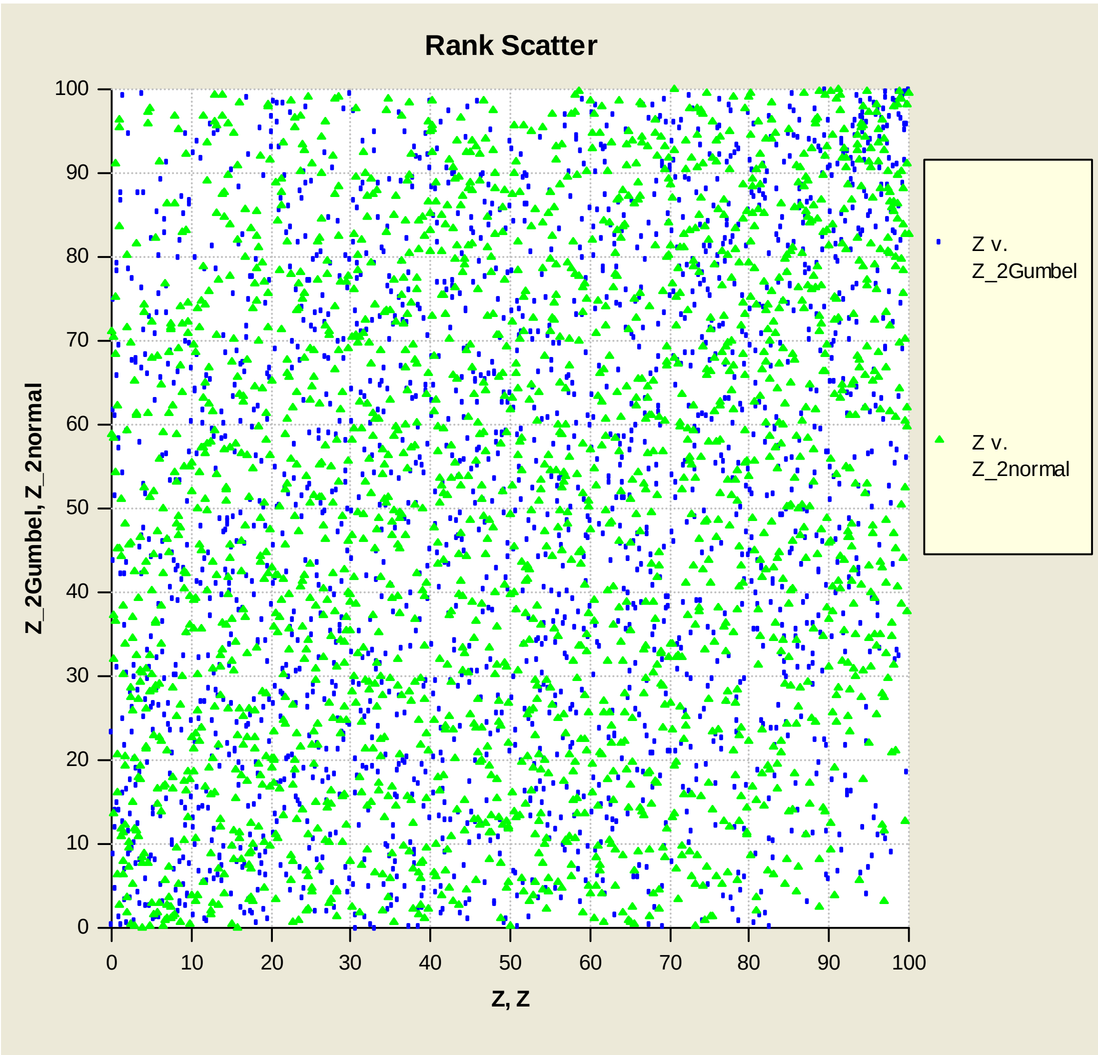


figure 3: Rank scatter plot of simulated Normal and Gumbel copulas.

The two scatter plots shown in Figure 3 show very similar patterns, both with a slightly lower density of points towards the upper left hand and lower right hand corner, and higher towards the other two corners. This indicates that the dependence structures simulated by the two copulas are very similar.

9. DISCUSSION OF RESULTS

The dependence between interest and inflation rates has a considerable impact on the distribution of the required reserve. The parameterization of the copulas in alternatives 3 and 4 require a sufficiently large history of reliable data, and one needs to assume that the dependence structure does not change over time. The approach in alternative 2 however provides an upper bound with regard to the dependence between the random errors of the two time series. Hence alternative 2 may be preferable if a prudent approach is sought and historic data are not considered sufficiently reliable.

The difference between the Normal and the Gumbel copula and the impact on the distribution of the required reserve is minimal. The Gumbel copula gives a better fit to the data than the Clayton copula. A fit of the Normal copula can not be shown in the same way as it does not belong to the family of so-called ‘Archimedean’ copulas.

Parameterization of an interest rate model based on historically observed rates may lead to results which are inconsistent with current market rates. Also, the use of a one-factor model can be regarded as too simplistic. However additional prudence can be built in by reducing the long term mean parameter b for example on the basis of projections by an economic forecasting bureau. The long term average interest rate parameter b of 6.7% appears high in the current environment, and leads to a continuous upward trend in the projected future interest rate.

10. STOCK MARKET RETURNS AND THE THREE DIMENSIONAL COPULA

In the preceding chapters we have limited ourselves to the estimation of a two-dimensional copula for the residuals of interest and inflation rates. In order to analyse the practical limitations of copulas for the modelling of the dependence of three, variables, we introduce a third variable: the return on the S&P 500 stock market index.

We will estimate the joint distribution of the residuals of interest and inflation rates and stock market returns. For practical purposes, this becomes relevant when then insurer decides to invest the assets backing the liabilities in a combination of stocks and bonds, rather than bonds only.

In the remainder of this chapter, we will first discuss the modelling of stock market returns over an extended period of time, and then the dependence between stock market returns, interest and inflation rates.

10.1 Parameterisation of a model for stock market returns

Define

$S(t)$: value of a common stock portfolio at the beginning of period t.
$M(t)$ return on the portfolio in monetary amounts during period t, assuming immediate reinvestment of dividends in the same stock portfolio.

As measure of return we use $R(t)$ defined as:

$R(\mathrm{t}) = (M(t) + S(\mathrm{t}))/S(\mathrm{t})$.

We will assume that

(1) $R(t)$ are independent and identically distributed for all t.

(2) $R(t)$ follow a lognormal distribution.

Assumption 1 reflects the weak form of the market efficiency hypothesis. Under this assumption stock market movements in the past can not be used be used to predict future movements, as the two are independent. There is some evidence against this hypothesis[11], and we will test the hypothesis using the AR(1) model described above.

Assumption 2 is also a widely used assumption for stock market returns, underlying for example the Black Scholes model. We will check the fit of the lognormal distribution against the observed observations by way of a QQ-plot and the previously used statistical tests.

10.1.1 Testing independence of stock market returns in consecutive periods

We use the AR(1) model specified in section 2 to test the hypothesis that stock market returns in consecutive years are independent by fitting the following model:

$$ln[R(t)] = a + b\ ln[R(t-1)] + \varepsilon(t),\quad t = 1,2,\ldots \qquad (3)$$

If the estimate of the coefficient b is not significantly different from 0, then the hypothesis of independence of $R(t)$ and $R(t+1)$ for all t is not rejected. This is in itself does not confirm independence of $R(t)$ and $R(t+1)$, as the assumption that b=0 only implies that $R(t)$ and $R(t+1)$ are uncorrelated, but not necessarily that they are also independent[12].

The parameterisation of *(3)* yields the following results:

Table 7: Parameter estimate AR(1) log stock market returns

	parameter estimate	standard error
a	0.064	
b	0.042	0.156
R^2	0.1%	

The parameter b is clearly not significant, and the model has no explanatory power considering the R^2 value of 0.1%.

[11] see Montier (2002)

[12] A well known example is the following: if $X \sim N(0,1)$ and $Y = X^2$, then $\rho(X,Y) = 0$ but obviously X and Y are not independent.

We will therefore assume returns are independent, and fit a lognormal distribution to the observed values of logarithms of stock market returns. The parameters of the fitted lognormal distribution are:

$\mu = 0.067$
$\sigma = 0.158$.

The QQ- plot and goodness-of-fit test results are as follows:

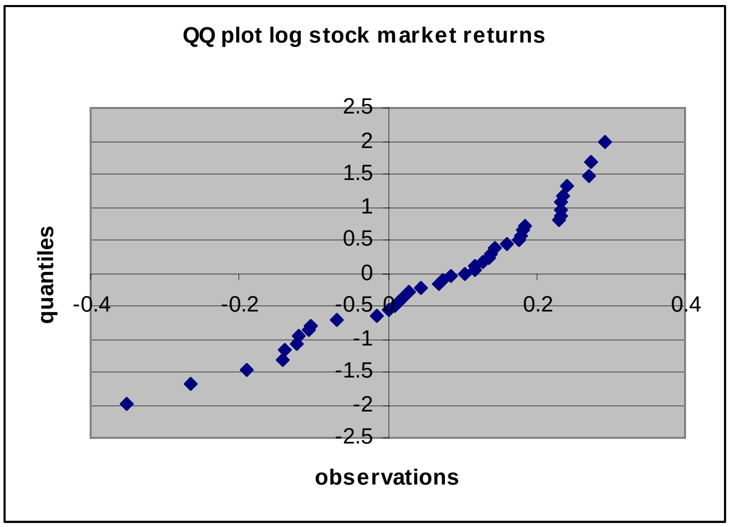


Table 8: Goodness-of fit test results (number of points: 41)

Test	**p-value**
Chi-squared	0.3014
Kolmogorov-Smirnov	>0.15
Anderson-Darling	0.05<=p<=0.1

Both the QQ-plot as well as the Anderson-Darling test statistic indicate that the fit of the distribution may not be good in the tails. This conclusion concurs with other research[13] that suggests stock market returns have heavier tails than the lognormal distribution.

We will nevertheless continue working on the assumption that the distribution of *R* is indeed lognormal.

[13] See Montier (2002)

10.2 Dependence between interest and inflation rates and stock market returns

Stock market returns, as well as interest rates and medical inflation are impacted by the same macro-economic factors. Therefore, dependence between stock market returns and the other two modelled variables can be expected.

As an initial exploration, we have estimated the linear correlation between the logarithm of stock market returns and the residuals of the interest and inflation rate models:

Table 9: linear correlation matrix (standard error of estimates between brackets)

Variable	$ln[R(t)]$	$\varepsilon_{l\_inf}(t)$	$\varepsilon_{int}(t)$
$Ln[R(t)]$	1		
$\varepsilon_{l\_inf}(t)$	-0.283(0.15)	1	
$\varepsilon_{int}(t)$	-0.240 (0.15)	0.20(0.15)	1

The negative correlation between $R(t)$ and $\varepsilon_{l\_inf}(t)$ is somewhat counterintuitive, as one would expect economic growth to drive on the one hand, stock market returns, and on the other, advances in medical technology giving rise to more expensive medical treatment. However, none of the estimated correlations are significantly different from 0.

We will now consider the following three dependence structures between $R(t)$, $\varepsilon_{l\_inf}(t)$ and $\varepsilon_{int}(t)$:

1. The three dimensional version of the Gumbel copula used above.
2. R(t) independent of $\varepsilon_{l\_inf}(t)$ *and* $\varepsilon_{int}(t)$*, Gumbel copula between* $\varepsilon_{l\_inf}(t)$ *and* $\varepsilon_{int}(t)$ *as before.*
3. Normal copula with three-dimensional correlation matrix estimated from the data.

1. The three dimensional Gumbel copula

The Gumbel copula can be extended to any number of variables. The three dimensional version is:

$$C(u,v,w) = \exp\{-[(-\ln u)^{\alpha} + (-\ln v)^{\alpha} + (-\ln w)^{\alpha}]^{1/\alpha}\}, \qquad \varphi(u) = (-\ln u)^{\alpha}.$$

A disadvantage of the Gumbel copula in more dimensions is that there is still only a single parameter that characterises the dependence between all pairs of individual variables. It is not possible to specify dependence between $R(t)$ and $\varepsilon_{l\_inf}(t)$ differently from the dependence between $\varepsilon_{l\_inf}(t)$ and $\varepsilon_{int}(t)$.

The graph below shows, for the dependence between $R(t)$ and $\varepsilon_{int}(t)$, the fit of the Gumbel copula with the previously fitted parameter ('Gumbel inf/int'), the optimal parameter for the dependence of $R(t)$ (Gumbel bestFit) and $\varepsilon_{int}(t)$, and the fit of the independent copula (Indpt).

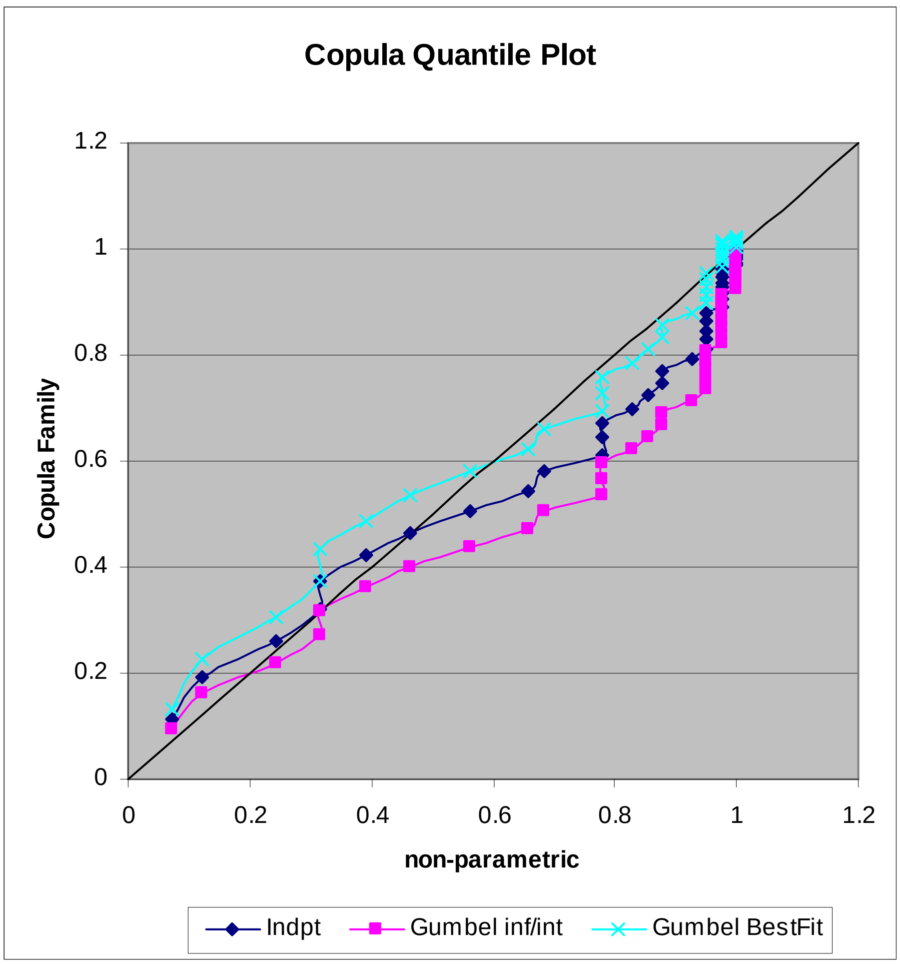


figure 4: Copula QQ-plot for *R(t) and* $\varepsilon_{int}(t)$.

Not surprisingly the previously parameterised Gumbel-copula does not fit very well.

The same plot for the dependence between $R(t)$ and $\varepsilon_{l\_inf}(t)$ is as follows:

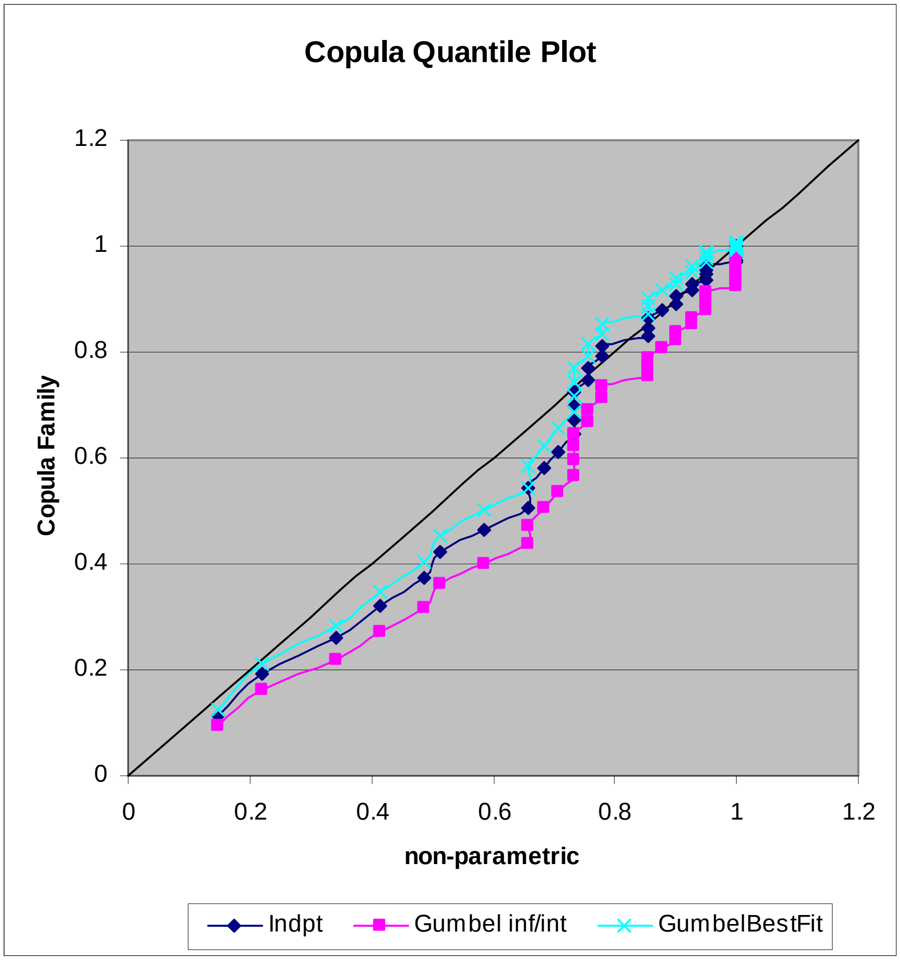


figure 5: Copula QQ-plot for *R(t)* and $\varepsilon_{l\_inf}(t)$**.**

Also here, we see that the optimal fit differs markedly from that obtained by the earlier estimated parameter. We will therefore consider the next two alternatives.

2. R(t) independent of $\varepsilon_{l\_inf}(t)$ *and* $\varepsilon_{int}(t)$.

Although the linear correlations with the other modelled variables are not significant, the QQ-plots above clearly suggests that there is dependence between *R(t)* and $\varepsilon_{int}(t)$. Nevertheless this alternative clearly provides a better fit than the Gumbel copula fitted to interest and inflation rates.

3. Using the Normal copula

The normal copula has the advantage that different correlations can be specified between each variable in a correlation matrix. Although the fit of the copula can not be inspected by way of a QQ-plot, we can compare the Gumbel and Normal copulas as was done in section 7. When applied to ln[$R(t)$], $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ which are all normally distributed, the Normal copula defines a multivariate normal distribution and thus makes a natural fit with the marginal distributions of the variables.

In the previous sections we have established that the fitted Normal copula for $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$ is very similar to the fitted Gumbel copula so that at least in this case there is no clear evidence that the Gumbel copula would be a better choice. Although none of the observed correlations between any of the three variables are significant, using the correlations estimated from the data provides a correlation structure closest to that present in the data.

We conclude that the normal copula is the best choice of the three alternatives considered.

10.3 Application to Present value of liabilities

Having defined the joint distribution of $R(t)$, $\varepsilon_{int}(t)$ and $\varepsilon_{l\_inf}(t)$, we can show the impact of various asset compositions on the distribution of the present value of the liabilities. Suppose the insurance company decides to invest that part of the provision that provides for the payment of the liability that falls due after five years in the S&P 500 index, and the first five years as before.

RR(0) as defined in section 3 now becomes:

$$RR(0) = \sum_{s=1}^{5}[Ac(s) \times \prod_{\tau=0}^{s-t-1} Df(t+\tau)] + \sum_{s=6}^{10}[Ac(s) \times \prod_{\tau=0}^{s-t-1} DfS(t+\tau)]$$

with $DfS(t) = \dfrac{1}{R(t)}$.

Using a normal copula with correlation matrix as in table 9, we can simulate the distribution of *RR(0)* (denoted as PV0 in the graph) to obtain the following result.

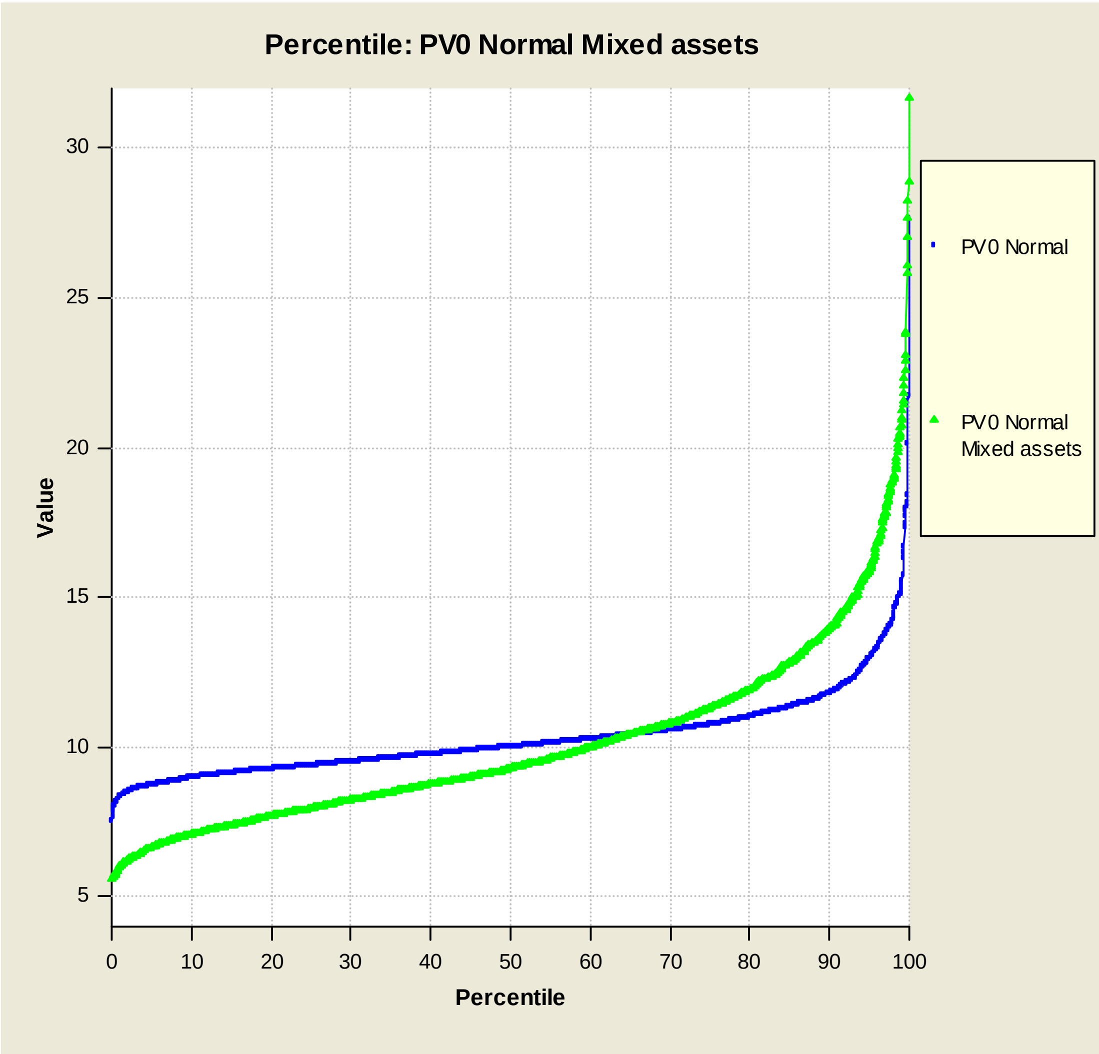


figure 6: Present value under the two different choices for asset composition

We can conclude that the present value of the liability when investing in the prescribed mix of common stock and bonds will be lower than in the case where all assets are bonds generating the 3 year interest rate, with a probability of approximately 65%.

11. CONCLUSION

The procedure described in chapter 7 works reasonably well as estimation method for the parameterisation and verification of goodness of fit of Archimedean copulas. A disadvantage of the Archimedean copula family is however that the dependence structure of each pair of variables within the copula is always identical. It is not possible to define different correlations between different variables within the copula.

The normal copula on the other hand does allow for the complete specification all pairwise dependencies by using a correlation matrix. Simulation of a normal copula using Choleski decomposition[14] is straightforward, and can be executed in many software packages, for example @Risk. An often cited drawback of the Normal copula however is the absence of so-called tail correlation, to allow for possibly higher correlation in the tails of the distribution. But even with data history going back more than forty years, dependencies in the extreme tail can not be parameterised on the basis of the actual data.

The parameterisation and goodness of fit verification of the various copulas can be further improved by applying maximum likelihood estimation of the parameters. This approach will allow for the determination of the standard error of the estimated parameters, and comparison of the goodness of fit of more than only Archimedean copulas.

Despite the disadvantages of copulas, such as the difficulties in parameterisation, using copulas for modelling dependencies has distinct advantages. It allows for the specification of a dependence structure for any combination of marginal variables in a sound and theoretically well-defined framework. This is a major improvement upon more traditional approaches such as the application of a correlation matrix to percentiles of marginal distributions.

[14] See Wang (1998)

APPENDIX

1. Medical Inflation rates

Area: *U.S. city*
Item: *Medical care*
Source:http://data.bls.gov/servlet/SurveyOutputServlet?data_tool=latest_numbers&series_id=CUUR0000SAM&output_view=pct_1mth

2. Interest rates

Rate of interest in money and capital markets
Federal Reserve System
Long-term or capital market
Government securities
Federal
Constant maturity
Three-year
Not seasonally adjusted
Twelve months ending December

Source: http://www.federalreserve.gov/releases/h15/data.htm#fn12

3. S&P 500

Source: finance.yahoo.com

S&P 500 adjusted index.. Adjustment refers to reinvestment of dividend into the index upon receipt thereof.

REFERENCES

[1] Box, George E, Gwilym M. Jenkins and Gregory C. Reinsel "Time Series Analysis" Prentice-Hall 1994.
[2] Butler, Cormac 'Mastering Value at Risk' Financial Times/Prentice Hall 1999.
[3] Cherubini, Umberto, Elisa Luciano and Walter Vecchiato "Copula Methods in Finance". Wiley Finance 2004.
[4] Denuit, Michel, and Jan Dhaene "Simple Characterizations of Comonotonicity and Countermonotonicity by Extremal Correlations" Catholic University Leuven 2003. http://www.econ.kuleuven.ac.be/tew/academic/actuawet/pdfs/Corel_Denuit_Dhaene_BAB.pdf
[5] Embrechts, Paul, Filip Lindskog, and Alexander McNeil "Modelling Dependence with Copulas and Applications to Risk Management" RiskLab September 2001.
[6] Embrechts, Paul, Alexander McNeil and Straumann 'Correlation and Dependence in Risk Management: Properties and Pitfalls' http://www.risklab.ch/ftp/papers/CorrelationPitfalls.pdf 1999.
[7] Frees, Edward W. and Emiliano A Valdez "Understanding Relationships Using Copulas" *North American Actuarial Journal*, 1998, 1-25.
[8] Kaufmann, Roger, Andreas Gadmer and Ralf Klett "Introduction to Dynamic Financial Analysis". Astin Bulletin, Vol. 31, No. 1 (May 2001) pages 213-249. http://www.casact.org/library/astin/vol31no1/213.pdf
[9] Mikosch, Thomas 'Copulas: Tales and Facts' 2005. http://www.math.ku.dk/~mikosch/Preprint/Copula/s.pdf
[10] Montier, James ' Behavioural Finance: Insights into irrational minds and markets' Wiley Finance 2002.
[11] Nelsen, Roger B "An Introduction to Copulas" Springer 1999.
[12] Wang, Shaun S "Aggregation of Correlated Risk Portfolios: Models and Algorithms" *PCAS* 1998, Vol. LXXXV, 848-939.