

# Copula Models of Economic Capital for Life Insurance Companies





# Copula Models of Economic Capital for Life Insurance Companies

**AUTHOR**

Arkady Shemyakin  
Huan Zhang  
Sydney Benson  
Regina Burroughs  
Jessica Mohr  
University of St. Thomas

**SPONSOR**

Joint Risk Management Research  
Committee

**Caveat and Disclaimer**

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information

Copyright © 2019 by the Society of Actuaries. All rights reserved.

# Copula Models of Economic Capital for Life Insurance Companies

Arkady Shemyakin, Huan Zhang,  
Sydney Benson, Regina Burroughs, and Jessica Mohr<sup>1</sup>

<sup>1</sup>University of St. Thomas, USA  
E-mail: a9shemyakin@stthomas.edu

## Abstract

The objective of this project is to construct, select, calibrate and validate a practically applicable copula internal model of economic capital for insurance companies. Copula methodology makes it possible to address multiple dependent risk factors. We identify the relevant set of asset and liability variables, and suggest a copula model for the joint distribution of these variables. Estimates of economic capital can be based on the tails of this joint distribution. Models are implemented in open source software (R and Microsoft EXCEL) and tested using simulated asset/liability data. The results are presented as a finished software product which can be utilized for customization and direct user application. The novelty of the approach consists in estimating interdependent mortality, morbidity, lapse and investment risks in one multivariate model. In particular, we address the challenges that life insurance companies face in the low interest environment. This approach requires a methodology of copula model comparison and selection and implementation of Monte Carlo simulation to the estimation of economic capital.

## 1 Project Objectives

Economic capital (EC) as a measure of the company's risk capital depends on a number of company-specific and external (both macroeconomic and industry-specific) risk variables and formulaic model constructs. The authors use a copula approach to build a predictive model estimating the amount of economic capital a life insurance company needs to protect itself against an adverse movement in interest rates, mortality, and other risk drivers. Predictive modeling requires a study of statistical dependence between diverse risks whose dependence can be expressed in terms of a joint distribution of risk variables using a copula function. Risk capital can then be estimated by value-at-risk (VaR) and tail value-at-risk (TVaR) values.

For example, under the European Union's Solvency II, the required capital is determined using a VaR at the 99.5% level; under the Swiss Solvency Test, the required

capital is determined using TVaR at the 99% level, see [9] for a comparison of the European Union's Solvency II system and the Swiss Solvency Test (SST). We will use both approaches at the 99.5% level for comparison purposes.

Economic capital modeling allows insurance companies to obtain better control of their risks and manage their business. The research paper sponsored by the SOA and Towers Perrin [13] outlines the general EC framework for a life insurance company and the major types of risk facing an insurance company. Mortality risk (including catastrophic, volatility, estimation, and trend risks) and interest rate risk are specified as key liability and asset related risks. Due to the nature of insurance products underwritten by a life insurance company, implementing EC for interest rate risk requires developing an advanced way of calculating capital needs for a mixture of mostly fixed-income securities with various maturities and credit qualities.

Many large insurers and subsidiaries of European insurers are using EC and stress testing models for solvency and high level capital planning considerations. However, there are many companies unfamiliar with implementing a model framework. One of the hurdles for new players is representing key dependent variables by their distributions, since industry and company experience is usually stated as a deterministic best estimate rate. Another major challenge is combining all modeled drivers of loss risk into a joint distribution. This is where the proper choice and use of a particular type of copula comes in [26], [29], [36].

Since the financial crisis of 2008-2009, the U.S. regulators adhere to the policy of maintaining interest rates at a very low level by historical standards. While this policy seems to produce a positive effect on overall economic development, it has created a challenging environment for the insurance industry. This effect is especially pronounced for life insurance companies struggling to gain investment returns that are sufficient for meeting their long term liabilities. Therefore our asset analysis is based on post-crisis years addressing the unique challenges of low interest markets.

Recently, the Federal Reserve has raised interest rates several times and signaled that it will continue raising rates in the nearest future. Higher interest rates would provide much needed relief to the life insurers that experienced a serious asset-liability mismatch during the last ten years (see, for example, [24]). However, there are several factors that might lead to preserving the low interest rate environment for the years to come. Among them is the sheer size of funds the Federal Reserve induced into the financial markets during its large-scale purchases of long-term government bonds and other securities (quantitative easing program) in 2008, 2010 and 2012. This has created a market where too many dollars are chasing too few investment opportunities. The other factor is rather modest long-term economic growth expectations coming from the Federal Reserve and shared by the market participants (see, for example, [22]).

Keeping this in mind, our research develops a practical approach to EC that would help a life insurance company calculate its capital needs in the low interest rate environment. In order to do so, we create an asset-liability framework modeling an insurance

company and use it to demonstrate how a copula-based approach can be applied to establishing economic capital that would assist the company in sustaining an extended period of low interest rates. The emphasis is made on statistical aspects of model selection: choice of particular copula type and marginal distributions of the asset and liability variables. Development of the model framework employed for the project uses traditional life products to illustrate copula utility. For products with interest sensitivity, equity market or other asset/liability linkages, use of the model will require possible expertise from the end user that may include further correlation factor study, marginal distribution development, use of R programming, etc. The severity of losses driving the ultimate economic capital levels in the report's illustrated base case is arbitrary though intended to be somewhat realistic per magnitude. As noted above, use of a low interest environment may also impact the severity of results. Section 2 describes the general aspects of model construction, while in Section 3 we discuss the details of model selection and calibration. Simulation from a copula and calculation of the end-of-the-year losses is covered in Section 4. Computer implementation and numerical examples are presented in Section 5, also covering some model assumptions and model validation aspects.

## 2 Model Construction

### 2.1 Internal Model for Economic Capital

Under modern insurance regulatory capital systems, like Solvency II, insurance companies have the option to employ an internal model for the calculation of their required capital. Here, we introduce an internal model which has been widely used in literature, e.g., [5], [7], and [12]. In line with the Solvency II Directive, basic own fund (BOF) is defined as the balance sheet amounts of assets over liabilities and subordinated liabilities. For simplification, we assume that the insurer does not hold any subordinated liabilities. Hence, the insurer's BOF, which is the net value of assets minus liabilities, at time 0 is

$$BOF^{(0)} = A^{(0)} - L^{(0)}, \quad (1)$$

where

$A^{(0)}$  represents the deterministic market value of the assets at time 0,

$L^{(0)}$  represents the deterministic value of the insurer's insurance liabilities at time 0, which is the insurer's total reserve at time 0.

Suppose the insurer makes investments into  $k$  assets, each with an annual return rate  $R_j$ ,  $j = 1, \dots, k$ . The portfolio weight of the  $j$ -th asset is  $w_j$ ,  $j = 1, \dots, k$ , satisfying  $\sum_{j=1}^k w_j = 1$ . Based on discrete compounding, the value of an insurer's assets at the end of the year can be expressed as:

$$A^{(T)} = A^{(0)}(1 + \mathbf{wR}),$$

where

$A^{(T)}$  represents the stochastic value of the insurer's assets at end of the year,

$\mathbf{w} = (w_1, \dots, w_k)$ , the vector of asset portfolio weights, which satisfies  $\sum_{j=1}^k w_j = 1$ ,

$\mathbf{R} = (R_1, \dots, R_k)$ , a random vector of asset returns.

The value of insurance liabilities at the end of the year depends on the insurer's annual excess loss rate, where excess loss represents the difference between the actual claim amount and expected claim amount. Analogous to the asset side, suppose the insurer has  $l$  lines of business, and each line of business has loss due to two causes. For example, the two causes for whole life insurance could be mortality and lapse. We define the annual excess loss rate for each business line and each cause as

$$X_{j,r} = \frac{C_{j,r} - E[C_{j,r}]}{L_j^{(0)}}, \quad (2)$$

where

$X_{j,r}$  represents the annual excess loss rate for the  $r$ -th cause of the  $j$ -th business line,  $C_{j,r}$  is a random variable representing the annual claim amount for the  $r$ -th cause of the  $j$ -th business line,

$L_j^{(0)}$  represents the reserve at the beginning of the year for the  $j$ -th business line.

The insurer's liabilities at the end of the year,  $L^{(T)}$ , is the sum of its total reserve at the end of the year and the annual total excess loss. Here, we assume that the total reserve at the beginning of the year is the same as the total reserve at the end of the year. This assumption is reasonable if the insurer has a stable liability portfolio and is in a stable external environment, e.g., interest rate and mortality do not change. Under this assumption,  $L^{(T)} - L^{(0)}$ , the change in the insurer's liabilities over one year, is just the annual total excess loss. Hence, the value of the insurer's liabilities at end of the year can be expressed as:

$$L^{(T)} = L^{(0)}(1 + \mathbf{v}\mathbf{X}),$$

where

$L^{(T)}$  represents the stochastic value of the insurer's liabilities at the end of the year, which is the sum of its total reserve at the end of the year and the annual total excess loss,

$\mathbf{X} = (X_{1,1}, X_{1,2}, X_{2,1}, X_{2,2}, \dots, X_{l,1}, X_{l,2})$ , a random vector of annual excess loss rate for each business line and each cause,

$\mathbf{v} = (v_1, v_1, v_2, v_2, \dots, v_l, v_l)$ , the vector of liability portfolio weights, where  $v_j = L_j^{(0)}/L^{(0)}$  denotes the weight of the  $j$ -th business line,  $j = 1, \dots, l$ .

Hence, the insurer's BOF at the end of the year is

$$BOF^{(T)} = A^{(T)} - L^{(T)}. \quad (3)$$

Then, the insurer's total loss during the one-year horizon can be expressed as

$$BOF^{(0)} - BOF^{(T)} = \left( A^{(0)} - L^{(0)} \right) - \left( A^{(T)} - L^{(T)} \right) = L^{(0)} \mathbf{v} \mathbf{X} - A^{(0)} \mathbf{w} \mathbf{R}.$$

The insurer's economic capital is the value at risk (VaR) of its total loss subject to a confidence level of 99.5% over a one-year period, that is,

$$EC = \text{VaR}_{99.5\%} \left( L^{(0)} \mathbf{v} \mathbf{X} - A^{(0)} \mathbf{w} \mathbf{R} \right), \quad (4)$$

which depends on the tails of the total loss distribution. However, due to a complex form of the joint distribution of  $(\mathbf{R}, \mathbf{X})$ , it will be difficult to obtain analytical expressions for the tails of the distribution of total loss. Therefore, we will perform Monte Carlo simulation of  $N$  independent copies of random vectors

$$(\mathbf{R}_i, \mathbf{X}_i), i = 1, \dots, N, \quad (5)$$

which can be used to estimate the tail probabilities with a threshold  $y$  as the direct count of tail points

$$\frac{1}{N} \sum_{i=1}^N I \left( L^{(0)} \mathbf{v} \mathbf{X}_i - A^{(0)} \mathbf{w} \mathbf{R}_i \leq y \right) \sim Pr \left( L^{(0)} \mathbf{v} \mathbf{X} - A^{(0)} \mathbf{w} \mathbf{R} \leq y \right). \quad (6)$$

Our model makes some simplifying assumptions with regard to the role of reserves and the use of total losses vs. excess losses. Though reserves that are already established for expected losses contain conservatism, we have assumed that the release of such reserves during the model horizon roughly matches the expected loss component of the total loss and will extinguish the expected loss one for one. As a result, we subtract the expected loss from the total loss to establish the amount of unexpected, or excess loss that is the key driver of an economic capital model. Our liability loss data is thus intended to represent excess loss only, by product and loss type.

Since a typical economic capital model horizon is one year in length, movement of reserves within that year are viewed as noise that may not impact the end of the year result. The exceptional case would be reserve changes between time 0 and time  $T$  that contemplate unexpected losses that are likely to occur after time  $T$  and need to be booked after time 0, but before the end of the year. In these cases, our recommendation is to scale up the baseline loss magnitude for time 0 to time  $T$  by incorporating this characteristic into the moments of the product's distribution or the estimated sample loss data set used. The end user is invited to explore general economic capital theory on this topic since the future losses driving this assumption do not occur in the time 0 to time  $T$  interval.

## 2.2 Modeling Asset Variables

We use a subset of Barclay indices to serve as proxies for our asset variables representing US bonds of different type and duration, emerging markets, mortgage backed

securities, and other investment instruments typical for the general account of a life insurance company. Such data can be observed on a monthly basis. Therefore, we will model asset variables as the percentage returns on certain indices  $P_j$ ,

$$R_j = \frac{P_j^{(T)} - P_j^{(0)}}{P_j^{(0)}}, \quad j = 1, \dots, k, \quad (7)$$

where  $P_j^{(0)}$  is the initial index value (beginning of the year) and  $P_j^{(T)}$  is the year-end value of the index.

Most of the fixed income indices are likely to exhibit autocorrelation and variable volatility. Therefore, we perform the conventional time series analysis as in [1] and [18] in order to isolate the stationary stochastic components  $U_j$ , filtering out major effects of autocorrelation (ARIMA). We have also considered GARCH filtering to address heteroskedasticity, but, while having an important effect for daily data, it proved to be statistically insignificant on a monthly basis. Then we develop a joint distribution model of  $\mathbf{U} = (U_1, \dots, U_k)$ , which will be used to model  $\mathbf{R} = (R_1, \dots, R_k)$ .

Let us consider times  $t = 1, 2, \dots, T$ , where the time increment will correspond to 1 month and  $T = 12$  for a full year. We will start with  $P_j^{(0)}$  and evaluate  $P_j^{(T)}$  in 12 time steps. Notice that this analysis can be done using end-of-the-month index values or average-of-the-month values, where the latter choice corresponds to the "Asian option" approach, which is becoming increasingly popular in insurance modeling. One step with  $t = 1, 2, \dots, T$  can be decomposed into substeps [1], [21], [23]:

- $P_j^{(t)} = P_j^{(t-1)} \exp(S_j^{(t)})$ , where  $S_j^{(t)}$  is the monthly log-return;
- $D_j^{(t)} = \Delta^d(S_j^{(t)})$ , where  $\Delta$  is the difference operator,  $\Delta f^{(t)} = f^{(t)} - f^{(t-1)}$ ,  $\Delta^d f^{(t)} = \Delta^{d-1} f^{(t)} - \Delta^{d-1} f^{(t-1)}$  reflecting possible non-stationarity;
- $D_j^{(t)} = \varphi_{j0} + \sum_{i=1}^p \varphi_i D_j^{(t-i)} + \sum_{i=1}^q \theta_i D_j^{(t-i)} + U_j^{(t)}$ , as in ARIMA( $p, d, q$ );  
In case of  $d = 0$ , no additional differences needs to be applied, and  $D_j^{(t)} = \Delta^0(S_j^{(t)}) = S_j^{(t)}$ .
- If residuals  $U_j^{(t)}$  expose clear heteroskedasticity (variable volatility), an additional GARCH step can be introduced to address this issue. So far we have not detected a necessity for this step in monthly data (see the more detailed discussion in Section 3.2).

Finally,  $U_j$  is a stochastic stationary component whose distribution belongs to one of a few popular parametric families. We considered the standard normal (Gaussian) distribution  $N(0, \sigma^2)$  or in the case of fatter tails and asymmetry, the skewed  $t$ -distribution



introduced in [2] (an alternative is provided in [18]) with density

$$d(x; \alpha, \nu, \xi, \omega) = 2 t_\nu(y) T_{\nu+1}(z), \quad (8)$$

where  $y = (x - \xi)/\omega$ ,  $z = \alpha y \sqrt{(\nu + 1)/(\nu + y^2)}$ , and

$$t_\nu(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\pi\nu}\Gamma\left(\frac{\nu+1}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Here  $t_\nu(y)$  and  $T_\nu(y)$  are the p.d.f. and the c.d.f. of a  $t$ -distribution with  $\nu$  degrees of freedom,  $\xi$  and  $\omega$  are the location and scale parameters,  $\alpha$  is the **skewness parameter**, and  $\Gamma(x)$  is the gamma function. The degree of freedom parameter can take on any real values such that  $\nu > 2$ , which guarantees the existence of variance. The skewness parameter can take on values from the interval  $(-1, 1)$ . The standard normal distribution can be treated as a degenerate case of skewed  $t$  with  $\nu \rightarrow \infty$  and  $\alpha = 0$ .

This model for any  $j = 1, \dots, k$  will be characterized by its own set of  $p + q + 5$  ARIMA and skewed  $t$ -distribution parameters (subscript  $j$  omitted):

$$(\varphi_0, \dots, \varphi_p, d, \theta_1, \dots, \theta_q, \alpha, \nu, \xi, \omega). \quad (9)$$

Estimation of model parameters can be done separately for ARIMA and  $U_j$  distribution fitting via MLE or Bayes estimation in the case when reasonable priors are available.

### 2.3 Modeling Liability Variables

A similar approach can be used to model liability variables if we obtain time series data sufficient for model calibration (estimation of parameters in (9)). In this study, in the absence of extensive time series liability data, we suggest a **simplified model**, which assumes a moving average (MA) monthly claim amount structure. We can then evaluate annual excess loss rates based on annual claim amount using (2).

Considering the  $r$ -th cause of loss in the  $j$ -th business line, as defined in section 2.1, let  $C_{j,r}$  be the annual claim amount, and  $C_{j,r}^{(t)}$ ,  $t = 1, \dots, 12$ , be the monthly claim amount satisfying

$$C_{j,r} = \sum_{t=1}^{12} C_{j,r}^{(t)}.$$

We assume that monthly claim amount  $C_{j,r}^{(t)}$ ,  $t = 1, \dots, 12$ , follows a modified moving-average (MA) of order 1 model, i.e.,

$$C_{j,r}^{(t)} = U_{j,r}^{(t)} + \beta_{j,r} U_{j,r}^{(t-1)} \quad (10)$$

where  $U_{j,r}^{(t)}$ ,  $t = 0, \dots, 12$ , are mutually independent and follow a common log-normal  $(\mu_{j,r}, \sigma_{j,r}^2)$  distribution. Hence, the annual claim amount is

$$C_{j,r} = \sum_{t=1}^{12} C_{j,r}^{(t)} = \beta_{j,r} U_{j,r}^{(0)} + (1 + \beta_{j,r}) \sum_{t=1}^{11} U_{j,r}^{(t)} + U_{j,r}^{(12)}. \quad (11)$$

Including a moving average term with parameter  $\beta_{j,r} > 0$  addresses the effect of month-to-month carryover of the claim amount. This model for any line of business  $j = 1, \dots, l$  and any cause of loss  $r = 1, 2$  would be characterized by its own set of moving average parameter  $\beta$  and 2 marginal distribution parameters (subscript  $j, r$  omitted):

$$(\beta, \mu, \sigma). \quad (12)$$

A widely used assumption is that the sum distribution of independent log-normal random variables is well approximated by another log-normal random variable, see [4], [6], and [33]. Hence, equations (10) and (11) imply that both monthly claim amount and annual claim amount approximately follow log-normal distributions, which is consistent with many empirical studies. Wilkinson's method, see [33], estimates parameters  $\mu_{j,r}$  and  $\sigma_{j,r}$  by matching the first and second moments, that is, by solving the following two equations:

$$\begin{aligned} E[C_{j,r}] &= 12(1 + \beta_{j,r}) e^{\mu_{j,r} + \sigma_{j,r}^2/2}, \\ \text{Var}(C_{j,r}) &= (\beta_{j,r}^2 + 11(1 + \beta_{j,r})^2 + 1) (e^{\sigma_{j,r}^2} - 1) e^{2\mu_{j,r} + \sigma_{j,r}^2}. \end{aligned}$$

In the presence of reliable data, one can assume a different autocorrelation structure on  $U_{j,r}^{(t)}$  and modify (10) accordingly. Also, instead of a log-normal model for  $U_{j,r}^{(t)}$ , one can consider Gamma, Weibull, or composite models which are made up by piecing together two weighted distributions at a specified threshold, see [8] and [34] for more details regarding fitting insurance liability data using composite models. In the case of composite models, the distribution of  $U_{j,r}^{(t)}$  is not preserved in  $C_{j,r}$  as defined in (11).

## 2.4 Copula Model of the Joint Distribution

Models for individual asset and liability variables considered in two previous subsections include distributions of the stationary stochastic components of vector  $\mathbf{V} = (U_1, \dots, U_k, U_{1,1}, \dots, U_{l,2})$ . The first  $k$  components, corresponding to assets, will be modeled by the skewed  $t$ -distribution and the last  $l \times 2$ , corresponding to liabilities, by the log-normal distribution. However, the components of this vector are not necessarily independent. In the case of asset variables, it is likely that the correlation between components is rather strong, and moreover, their association may go beyond the correlation, getting stronger in the tails. Tail dependence may be successfully modeled by copulas [14]. In the case of some liability variables (mortality and morbidity related losses), we may assume independence from the market and, therefore, from the asset variables. However, for lapse related losses we allow for a certain dependence on the market. Thus, we consider a general copula model for all  $k + l \times 2$  variables.

Assume that  $F_j(u)$  is the c.d.f. of the  $j$ -th component  $U_j$  of the vector  $\mathbf{U}$ . Then a copula model is defined for the joint distribution of  $\mathbf{U}$  as

$$C_\gamma(F_1(u_1), \dots, F_{k+2l}(u_{k+2l})) = P(U_1 \leq u_1, \dots, U_{k+l \times 2} \leq u_{k+l \times 2}), \quad (13)$$

where  $C_\gamma$  is a special *copula function* with the set of parameters  $\gamma$ . We consider copulas of two types.

**Elliptical copulas.** The elliptical distribution  $Q_{d,R}$  of a random vector  $\mathbf{t} = (t_1, \dots, t_d)$  can be defined by its joint density function

$$|\Sigma|^{-1/2} k((t - \mu)^T \Sigma^{-1} (t - \mu)),$$

where  $\mu$  is a  $d \times 1$  vector of means,  $\Sigma$  is a positively defined  $d \times d$  covariance matrix, and  $k(x)$  is some non-negative function of one variable integrable over the entire real line. Matrix  $R$  with elements  $R_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii} \times \Sigma_{jj}}$  is the correlation matrix determining all pairwise associations between the components of the random vector  $\mathbf{t}$ . Define also by  $Q_i(t_i)$  the marginal distribution of  $t_i$ . Then we can define an elliptical copula as

$$C_R(t_1, \dots, t_d) = Q_{d,R}(Q_1^{-1}(t_1), \dots, Q_d^{-1}(t_d)). \quad (14)$$

The most popular elliptical copula is the Gaussian copula, which, combined with marginal distributions  $t_i = F_i(u_i)$  of the data vector  $\mathbf{U} = (U_1, \dots, U_d)$ , defines the joint distribution  $H(\mathbf{u})$  of vector  $\mathbf{U}$  as

$$H(\mathbf{u}) = C_R(t_1, \dots, t_d) = \Phi_{d,R}(\Phi^{-1}(F_1(u_1)), \dots, \Phi^{-1}(F_d(u_d))), \quad (15)$$

where  $\Phi(u)$  is the standard normal distribution and  $\Phi_{d,R}$  is  $d$ -variate normal with zero mean, unit variances and correlation matrix  $R$ . Off-diagonal elements of matrix  $R$  describe pairwise associations, so the strength of the association may differ for different pairs of components of vector  $\mathbf{U}$ . Gaussian copulas have been widely used in finance, but proved to be inefficient in estimating the tail dependence [25].

Another popular choice of an elliptical copula is Student's  $t$ -copula [10],

$$H(\mathbf{u}) = C_R(t_1, \dots, t_d) = T_{d,\eta,R}(T_\eta^{-1}(F_1(u_1)), \dots, T_\eta^{-1}(F_d(u_d))), \quad (16)$$

where  $T_\eta(x)$  is a  $t$ -distribution with  $\eta$  degrees of freedom, and  $T_{d,\eta,R}$  is  $d$ -variate  $t$ -distribution with  $\eta$  degrees of freedom, and correlation matrix  $R$ . In our case dimensionality is  $d = k + 2l$ . It works sufficiently well for financial applications, which encounter both lower and upper tail dependence. An attractive feature of the  $t$ -copulas is a possibility to treat the correlation parameters, defined by the correlation matrix, and the degree of freedom parameter separately [36]. The total number of copula parameters is  $(k + 2l)(k + 2l - 1)/2$  distinct terms of the symmetric correlation matrix  $R$  and one number of degrees of freedom  $\eta$ .

**Archimedean copulas.** Clayton's, Frank's and Gumbel-Hougaard's copula families provide a valid alternative to the use of elliptical copulas for modeling joint distribution tails. However, in dimensions higher than  $d = 2$  they require additional definition of the hierarchical structure (vines and nested copulas being two main options, see [35]). We will consider Clayton's nested copula family as in [20], [30] along with Student's  $t$ -copula as possible choices. Clayton's nested copula is built using a stepwise hierarchical procedure, such as

$$C_{\gamma_{1,2}}(t_1, t_2, t_3) = C_{\gamma_2}[C_{\gamma_1}(t_1, t_2), t_3] \quad (17)$$

from Clayton’s pair copulas

$$C_\gamma(t_1, t_2) = \max\{(t_1^{-\gamma} + t_2^{-\gamma} - 1)^{-1/\gamma}, 0\}, \gamma \in [-1; 0) \cup (0; \infty). \quad (18)$$

The total number of parameters for a fully nested Archimedean copula is  $k + l - 1$ , and could be higher for partially nested models and vines [32].

### 3 Model Selection and Calibration

#### 3.1 Variable Selection

Since 1973, the Bloomberg Barclays Indices [3] have been the most widely used indices for fixed income investors looking for representative benchmarks to measure asset-class risks and returns. On August 24, 2016, Bloomberg acquired these assets from Barclays. Barclays and Bloomberg have partnered to co-brand the indices as the Bloomberg Barclays Indices for an initial term of up to five years. The University of St. Thomas provides access to the Bloomberg terminal and index data are available through this for our research. We retrieved daily settlement data for the last 15 years in order to collect information for our study and model calibration. We used monthly averages of daily settlement prices. In order to replicate a typical asset portfolio of a life insurance company, see [37], we selected nine ( $k = 9$ ) Bloomberg Barclay indices (see Table 1).

Table 1: Selected Bloomberg Barclay indices.

Number	Ticker	Index
1	LUMS(TRUU)	Barclays US MBS Index
2	LUAC(TRUU)	Barclays Capital U.S. Corporate Bond Index
3	EMUS(TRUU)	Barclays Capital Emerging Markets Bond Index
4	LUCM(TRUU)	Barclays Capital U.S. CMBS (ERISA Only) Index
5	LU35(TRUU)	US Aggregate 3-5 Years
6	LU57(TRUU)	US Aggregate 5-7 Years
7	LU71(TRUU)	US Aggregate 7-10 Years
8	LF98(TRUU)	Barclays Capital U.S. Corporate High Yield Bond Index
9	LU13(TRUU)	Barclays Capital U.S. 1-5 Year Corporate Bond Index

Indices 2, 5-7, and 9 correspond to U.S. corporate bonds of various duration, indices 1 and 4 represent mortgage backed securities (both home and commercial mortgages), index 3 represents emerging markets, and index 8 represents high yield bonds. The mix of these nine indices (with variable weights) can represent a wide range of investment strategies used by life insurance companies. We used daily data for preliminary analysis, and then applied end-of-the-month and monthly averages for time series model development. A monthly periodicity makes it possible to synchronize asset and

liability data, since the latter are not likely to be observed on a daily basis. We selected monthly averages (similar to Asian option approach) as exhibiting the most stable patterns to implement the models described in Section 2.2 and further copula modeling as introduced in Section 2.4. The results of the preliminary estimation are summarized in Section 3.2.

For liability modeling we selected six variables representing losses characterizing a typical product mix of a life insurance company.

Table 2: Selected loss variables.

Number	Product	Source of Loss
1	Term Life (TL)	Mortality
2		Lapse
3	Whole Life (WL)	Mortality
4		Lapse
5	Disability Income (DI)	Morbidity
6		Lapse

At this point we do not have reliable monthly time series data that would allow us to obtain empirical estimates of the parametric distributions for liability variables. The development of our sample data for liability distribution analysis is the best estimate, based upon the judgement of an actuary familiar with the concept of economic capital and is intended to be illustrative and not necessarily indicative of any type or size of the company. Our process of the distribution elicitation roughly follows this procedure:

- Start with Solvency II guidance with regard to moments and extremes of loss.
- Use historical/anecdotal evidence from the last 200 years; companies may use 200 years if following the most standard approach. An example might be the Spanish flu outbreak of 1918 or plague outbreaks of middle age times.
- Form an industry best practice view of outlier losses by networking or using an expert who has seen a number of industry models in use. Our actuarial expert has seen about 5 models and their driving assumptions (and was careful not to disclose any confidential information or reuse it).
- Select mean, standard deviation and 99.5% points for each loss distribution by product and loss type.
- Using the above and best estimation, develop the distribution model and the estimates of its mean and variance that seem logical compared to experience with year to year losses on the job as well as the points selected in the previous step.

Though this approach does not produce actual data used in a consulting or company model, we feel that this procedure gives us a usable starting point for liability loss drivers. It is also key to stress that the purpose of this project is not to improve robustness of liability loss assumption development, but to demonstrate the use of copulas

for economic capital modeling.

One of the further objectives of the project is to make the end users of our product able to enter their own liability data into the model template. With this approach, if the end users obtain an estimate of their annual expected claim amount and its standard deviation for all business lines along with an estimate of reserve, they can develop a model for their particular product mix. The input data we use in the development of EC calculations are summarized in the attached file *EC Inputs and Results.xlsm*.

### 3.2 Parametric Estimation for Asset Variables

**ARIMA Model.** We begin with an ARIMA analysis of the asset variables (nine indices specified in Table 1). All procedures are implemented in *R* using `auto.arima` procedure from the *forecast* package and the package *rugarch* [15], which is convenient for diagnostics of heteroskedasticity in case a GARCH component is to be added in the future. We carried out the time series analysis of daily settlement prices as well as monthly averages obtaining normalized residuals  $U_j$ ,  $j = 1, \dots, k$ . The analysis of monthly averages suggests optimal values for ARIMA parameters  $p, d, q$ . The transformation to log-returns in the construction of Section 2.2 achieves stationarity, thus further ARIMA differencing appears to be rarely necessary; for eight out of nine indices  $d = 0$ . A GARCH(1,1) step does not significantly improve overall performance of the models for monthly averages since the effect of variable volatility is much more pronounced in the case of daily values. This can be demonstrated in Figure 1 and Figure 2, where the time series plot of ARIMA residuals is graphically compared with the ARIMA/GARCH residuals for the index LUMS(TRUU).

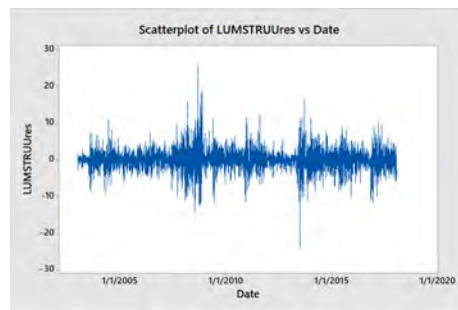


Figure 1: LUMS(TRUU) ARIMA residuals, daily values.

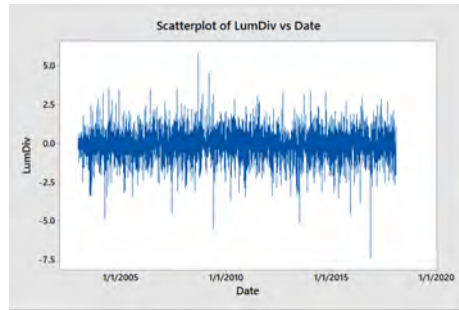


Figure 2: LUMS(TRUU) ARIMA/GARCH(1,1) Residuals, Daily Values.

In Figures 3 and 4 we can compare ARIMA residuals  $V^{(t)}$  with ARIMA/GARCH residuals  $U^{(t)}$  for the same index, LUMS(TRUU). Graphically, the difference between these two is not clearly expressed. Table 3 contains the values of ARIMA coefficients as estimated via MLE [15]. Detailed results of parametric estimation are provided in the worksheet *ARIMA Parameters* of the attached file *EC Inputs and Results.xlsm*.

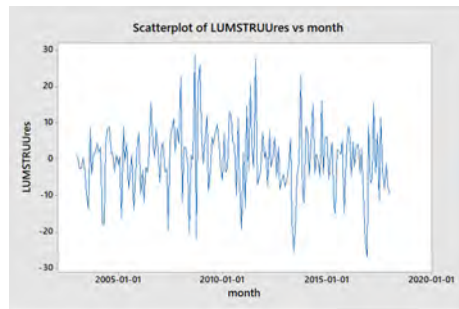


Figure 3: LUMS(TRUU) ARIMA residuals, monthly averages.

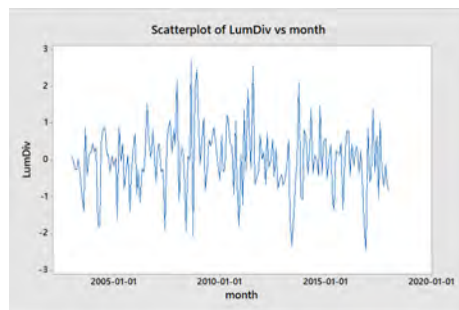


Figure 4: LUMS(TRUU) ARIMA/GARCH(1,1) residuals, monthly averages.

Table 3: Estimation of ARIMA parameters for asset variables

Index	ARIMA	$\varphi_0$	$\varphi_1$	$\varphi_2$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
LUMS	(2,0,2)	.009	-1.35	-.37	1.62	.66	0	0
LUAC	(1,0,3)	.009	-.97	0	1.42	.30	-.15	0
EMUS	(0,0,1)	.007	0	0	.52	0	0	0
LUCM	(0,0,4)	0	0	0	.60	.04	.22	.40
LU35	(2,0,2)	.008	-1.33	-.35	1.62	.65	0	0
LU57	(1,0,3)	.007	-.98	0	1.24	.07	-.21	0
LU71	(2,0,3)	.001	-.88	.09	1.17	-.07	-.29	0
LF98	(2,0,3)	.015	-.39	-.74	.82	1.00	.18	0
LU13	(1,1,3)	0	-.98	0	0.38	-.84	-.32	0

**Marginal Distributions.** Goodness-of-fit analysis was performed for the residuals  $U_j$ , and the  $p$ -values for both the normal and skewed  $t$ -distribution were calculated. It became clear that while the normal distribution provided a reasonable fit for most indices, the most interesting ones (EMUS related to emerging markets, LUCM related to commercial mortgage backed securities, and LF98 for high yield bonds) exhibit tails fatter than normal and require the use of the skewed  $t$ -distribution. Parametric estimates and goodness-of-fit results are shown in Table 4 and summarized in the worksheet *Asset Parameters* of the attached file *EC Inputs and Results.xlsm*.

Table 4: Estimation of marginal distributions for asset variables

Index	$\alpha$	$\nu$	$\omega$	$\xi$	GOF skewed $t$	GOF normal
LUMS	-.197	5.590	.005	.001	0.86	0.35
LUAC	-.593	4.219	.010	.005	0.82	0.14
EMUS	-.320	3.236	.012	.004	0.98	0.002
LUCM	.036	2.335	.008	.002	0.77	0.003
LU35	-.706	8.533	.005	.003	0.96	0.74
LU57	-.525	4.817	.007	.003	0.97	0.57
LU71	-.576	3.964	.009	.004	0.98	0.15
LF98	.007	2.769	.012	0	0.99	0.008
LU13	.235	5.915	.002	-.001	1.00	0.67

### 3.3 Parametric Estimation for Liability Variables

For the simplified liability model (11) suggested in Section 2.3 for  $U_{j,r}$ ;  $j = 1, \dots, 3$ ;  $r = 1, 2$ , we used the 200 sample years of experience to check the goodness-of-fit of log-normal, Gamma, and Weibull distributions, which are satisfactory. However, composite distributions (e.g., log-normal/Pareto) demonstrated slightly better fit of the



tail values. AIC (Akaike Information Criterion) defined as

$$AIC = 2m - 2 \ln L,$$

where  $m$  is the total number of parameters and  $L$  is the maximum of the likelihood function for a given model, was used for model comparison. The log-normal model was chosen due to the simplicity of implementation of the moving average model (11).

The results of the parametric estimation in the log-normal model are demonstrated in Table 5, where parameters  $\mu$  and  $\sigma$  correspond to monthly values  $U_{j,r}$ ,  $j = 1, \dots, 3$ ;  $r = 1, 2$  obtained directly from estimates  $E(C_{j,r})$  and  $Var(C_{j,r})$  under the assumptions of  $\beta_{j,r} \equiv 0.1; 0.5; 0.99$ ;  $j = 1, \dots, 3$ ;  $r = 1, 2$ . The approach of Section 3.1 was used as suggested by industry consultants according to the Solvency II recommendation. Different values of  $\beta_{j,r}$  lead to slightly different values for  $\mu$  and  $\sigma$ . Detailed inputs for the estimation of  $\mu$  and  $\sigma$  with and without the additional 1.8 factor applied to the estimate of the standard deviation of claim amount  $C_{j,r}$ , as recommended by industry consultants, are provided in the worksheet **Liability Inputs** of the attached file **EC Inputs and Outputs.xlsx** and the code necessary for this estimation is included in **EC Simulation.R**.

Table 5: Estimation of marginal distributions for liability variables

Liability	$\beta = .1$		$\beta = .5$		$\beta = .99$	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
TL Mortality	16.36	.183	16.05	.185	15.76	.186
TL Lapse	14.99	.045	14.67	.046	14.39	.046
WL Mortality	14.61	.183	14.30	.185	14.02	.186
WL Lapse	13.24	.045	12.93	.046	12.65	.046
DI Morbidity	16.37	.217	16.06	.220	15.77	.220
DI Lapse	14.78	.045	14.47	.046	14.19	.046

### 3.4 Choice and Calibration of Copula

We apply copula models to the normalized residuals of the asset variables  $U_j$ ,  $j = 1, \dots, k$  without a GARCH correction (GARCH(0,0)). The choice of a copula model and its calibration (parameter estimation) is carried out based on the data for asset variables, for which simultaneous time series are available. Liability variables are added to the set at the next stage involving the simulation from a copula and the estimation of economic capital (see Section 4). In order to compare the performance of Gaussian copulas,  $t$ -copulas, and Archimedean copula structures, we will follow the model selection procedures outlined in [17], [21], and [23] for stock index data. At this point, the AIC analysis in Table 6 (the lower AIC value corresponds to the better model) suggests that, for the restricted set of variables  $j = 1, \dots, k$ , the  $t$ -copula outperforms the other two classes, which is consistent with the literature, see [36]. We use the  $t$ -copula

as the model of choice in the current setting. However, it is possible that a more accurate choice of the Archimedean copula and its hierarchical structure may suggest a different model selection, when liability variables are included in the study. For comparison purposes, in Figure 5 we include a nesting diagram for the Clayton’s copula, which provides some insight on relative distances between the asset classes.

Table 6: Copula model comparison

Copula	Marginals	AIC
Normal	Normal	-12775
	Skewed $t$	-12381
$t$	Normal	-12847
	Skewed $t$	<b>-12984</b>
Clayton	Normal	-10814
	Skewed $t$	-11556

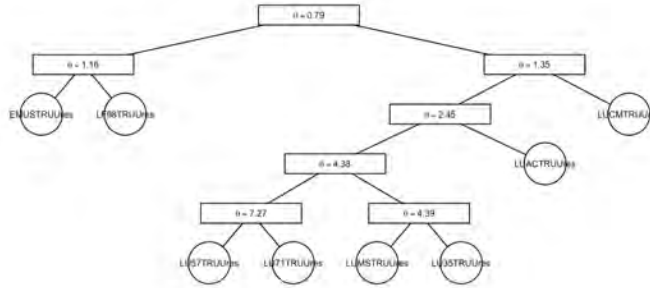


Figure 5: Nesting structure for the hierarchical Clayton copula.

Table 7 shows the estimated correlation parameters of the  $t$ -copula forming a symmetric  $9 \times 9$  correlation matrix  $R_1$ . The estimate of the copula degree of freedom parameter is  $\eta = 2.4155$ , which, in the model, is the common value for all variables. In this work, the R packages *sn* [2] and *copula* [19] were applied, with a combination of generalized method of moments and MLE used to estimate copula parameters. More detailed results of parametric estimation and model selection are provided in the worksheet *Correlation Table* of the attached file *EC Inputs and Results.xlsx*.

Table 7: Estimated correlations for  $t$ -copula:  $R_1$ , lower triangular view.

Ticker/Number	1	2	3	4	5	6	7	8	9
LUMS	1								
LUAC	.61	1							
EMUS	.46	.81	1						
LUCM	.38	.50	.51	1					
LU35	.91	.75	.57	.48	1				
LU57	.86	.76	.58	.58	.95	1			
LU71	.85	.79	.57	.56	.90	.97	1		
LF98	.16	.60	.77	.47	.26	.29	.24	1	
LU13	.73	.63	.49	.38	.87	.76	.69	.24	1

## 4 Simulation

### 4.1 Simulation from Copula

We concentrate on the  $t$ -copula suggested by the model comparison in Table 6 as the best choice for the set of the asset variables. Applying the copula model to all 15 variables (including liabilities) will allow us to generate month-by-month values of ARIMA innovations for assets  $U_j, j = 1, \dots, k = 9$ , and MA innovations for liabilities  $U_{j,r}, j = 1, \dots, 3; l = 1, 2$ . The values in Section 4.2 are used to generate multiple future scenarios of the end-of-the-year development of our asset/liability portfolio and then, in Section 4.3, are applied to the estimation of economic capital. If all copula parameters (including the parameters of the marginal distributions) are defined, we can use R packages [31] to perform this simulation.

For each asset variable  $j = 1, \dots, k = 9$  we have defined the set of parameters, which we can use in the  $t$ -copula framework:  $(\alpha_j, \nu_j, \xi_j, \omega_j)$ . For each  $j = 1, \dots, 3; l = 1, 2$  we have defined the set of parameters, which we can use in the  $t$ -copula framework:  $(\mu_{j,r}, \sigma_{j,r})$ . Copula parameters  $\eta$  (copula degrees of freedom) and symmetric correlation matrix  $R$ , size  $15 \times 15$ , later defined in Table 10, are also necessary for the simulation. However, only part of the matrix  $R_1$  (asset correlations) is already estimated. We need additional information on liability-to-liability correlations and asset-to-liability correlations. The nature of the data we use to estimate liability variables does not provide enough information for such estimation. Hence we need to elicit expert information to supplement our estimates from the data.

In order to establish preliminary estimates of the correlations between asset and liability variables, we will use Table 8 based on our analysis of Solvency II requirements and expert estimates provided by industry consultants. These values express the idea that though it is not clear whether there exists an association between mortality and morbidity on one hand, and the state of the market on the other hand, we can safely assume that losses due to lapses will be higher at the lower markets.

Table 8: Correlations of asset and liability variables:  $R_2$ .

Liability/Index	1	2	3	4	5	6	7	8	9
TL Mortality(10)	0	0	0	0	0	0	0	0	0
TL Lapse(11)	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1
WL Mortality(12)	0	0	0	0	0	0	0	0	0
WL Lapse(13)	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1
DI Morbidity(14)	0	0	0	0	0	0	0	0	0
DI Lapse(15)	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1

Similarly, expert opinion based on the Solvency II requirements suggests the correlation structure of liability variables summarized in Table 9. Combining symmetric matrices  $R_1$  from Table 7 and  $R_3$  from Table 9 with the  $6 \times 9$  block matrix  $R_2$  from Table 8, we construct the  $15 \times 15$  correlation matrix as suggested in Table 10.

Table 9: Correlations of liability variables:  $R_3$ , lower triangular view.

Liability/Index	1	2	3	4	5	6
TL Mortality (10)	1					
TL Lapse (11)	.3	1				
WL Mortality(12)	.5	.1	1			
WL Lapse(13)	.1	.5	.3	1		
DI Morbidity(14)	.25	.1	.25	.1	1	
DI Lapse(15)	.1	.1	.1	.1	.3	1

Table 10: Structure of correlation matrix  $R$ .

Variable	1-9	10-15
1-9	$R_1$	$R_2^T$
10-15	$R_2$	$R_3$

The attached file, *EC Inputs and Results.xlsm*, summarizes all parametric estimates needed to perform the simulation from the  $t$ -copula and further year-end value calculations. We use the R package *sn* [2] to simulate values from the skewed  $t$ -distribution. The corresponding R code is attached as *EC Simulation.R* and an example of the simulation result (1000 years) is provided in *EC Inputs and Results.xlsm*

## 4.2 Calculation of Year-end Values

**Results.** We obtain  $N$  independent copies of the end-of-the-year values of portfolio variables as relative (percentage) investment gains and excess losses

$$(\mathbf{R}_i, \mathbf{X}_i), i = 1, \dots, N, \quad (19)$$

which will be used in the final calculation of economic capital.

**Inputs.** We begin with the results of the simulation in Section 4.1: for each month  $t = 1, \dots, 12$  we define a  $k + l \times 2 = 15$ -dimensional vector

$$\mathbf{u}^{(t)} = (u_1^{(t)}, \dots, u_9^{(t)}, u_{1,1}^{(t)}, \dots, u_{3,2}^{(t)}), \quad (20)$$

where  $k = 9$  is the number of asset variables and  $3 \times 2 = 6$  is the number of liability variables. In this array of  $15 \times 12$ , all twelve columns representing months of one year are independent, and in each column the dependence between the rows corresponds to the  $t$ -copula structure.

Additional inputs are provided by the parameters of the ARIMA  $(p_j, d_j, q_j)$  models estimated for all asset variables  $j = 1, \dots, k = 9$  as

- autoregressive parameters  $\varphi_{j0}, \varphi_{j1}, \dots, \varphi_{jp_j}$ , and
- moving average parameters  $\theta_{j1}, \dots, \theta_{jq_j}$ .

For all  $j$  it holds that  $p_j \leq 2$  and  $q_j \leq 4$ , so that we can consider a rectangular array of parameters size  $(\max(p_j) + 1) + \max(q_j) \times j = 7 \times 9$ , substituting zeros where needed. Notice also that the differencing (integrated) parameter  $d_j = 0$  for  $j = 1, \dots, 8$  and  $d_9 = 1$  for the index LU13(TRUU). Therefore we address this index separately.

As the initial values for the ARIMA steps, we are required to obtain for asset variables  $j = 1, \dots, k = 9$ , the values  $P_j^{(0)}, P_j^{(-1)}, P_j^{(-2)}$ . In the case of an absence of reliable information, we can use only  $P_j^{(0)}$  and make a simplifying assumption  $P_j^{(-2)} = P_j^{(0)}$  and  $P_j^{(-1)} = P_j^{(0)}$ .

**Assets.** For the asset variables, our goal is to evaluate the percentage returns on certain indices  $P_j$ ,

$$R_j = \frac{P_j^{(T)} - P_j^{(0)}}{P_j^{(0)}} = \frac{P_j^{(T)}}{P_j^{(0)}} - 1, j = 1, \dots, k = 9, \quad (21)$$

where  $P_j^{(0)}$  is the initial index value (beginning of the year) and  $P_j^{(T)}$  is the year-end value of the index. Notice that  $ARIMA(p, d, q)$  models are suggested not for index values directly, but for for monthly log-returns

$$S_j^{(t)} = \ln \frac{P_j^{(t)}}{P_j^{(t-1)}}, \quad (22)$$

defined for all values of  $t = 1, \dots, T = 12$ , and also if needed for  $t = 0, -1, -2$ , etc. Thus equation (3) can be rewritten for  $j = 1, \dots, 9$  as

$$R_j = \exp\left(\sum_{t=1}^{12} S_j^{(t)}\right) - 1 \quad (23)$$

and for LU13(TRUU) in the presence of differencing as

$$R_j = \exp\left(12S_j^{(0)} + \sum_{t=1}^{12}(12-t+1)D_j^{(t)}\right) - 1, \quad (24)$$

where  $D_j^{(t)} = S_j^{(t)} - S_j^{(t-1)}$  is the first difference operator. The calculation of  $S_j^{(t)}$  and  $D_j^{(t)}$ , to be plugged into (23) and (24), is provided in the **ARIMA** subsection below.

### ARIMA

Case 1 ( $j = 1, \dots, 8$ , input for formula (23)): Let us consider times  $t = 1, 2, \dots, T$ , where the time increment will correspond to 1 month and  $T = 12$  for a full year. We will start with initial values  $S_j^{(0)}$  and  $S_j^{(-1)}$ , and evaluate  $S_j^{(t)}$  for all 12 time steps. If the initial values are not available, in accordance with the simplifying assumption above, take  $S_j^{(0)} = S_j^{(-1)} = 0$ . Initial values used in the EC calculation below are provided in worksheet **Initial ARIMA** of the file **EC Inputs and Results.xlsm**.

The algorithm is based on the ARIMA(p,0,q) model:

$$S_j^{(t)} = \varphi_{j0} + \sum_{i=1}^2 \varphi_{ji} S_j^{(t-i)} + u_j^{(t)} + \sum_{i=1}^4 \theta_{ji} u_j^{(t-i)}. \quad (25)$$

We will implement 12 steps:

- $S_j^{(1)} = \varphi_{j0} + \varphi_{j1} S_j^{(0)} + \varphi_{j2} S_j^{(-1)} + u_j^{(1)}$
- $S_j^{(2)} = \varphi_{j0} + \varphi_{j1} S_j^{(1)} + \varphi_{j2} S_j^{(0)} + u_j^{(2)} + \theta_{j1} u_j^{(1)}$
- $S_j^{(3)} = \varphi_{j0} + \varphi_{j1} S_j^{(2)} + \varphi_{j2} S_j^{(1)} + u_j^{(3)} + \theta_{j1} u_j^{(2)} + \theta_{j2} u_j^{(1)}$
- $S_j^{(4)} = \varphi_{j0} + \varphi_{j1} S_j^{(3)} + \varphi_{j2} S_j^{(2)} + u_j^{(4)} + \theta_{j1} u_j^{(3)} + \theta_{j2} u_j^{(2)} + \theta_{j3} u_j^{(1)}$
- $S_j^{(5)} = \varphi_{j0} + \varphi_{j1} S_j^{(4)} + \varphi_{j2} S_j^{(3)} + u_j^{(5)} + \theta_{j1} u_j^{(4)} + \theta_{j2} u_j^{(3)} + \theta_{j3} u_j^{(2)} + \theta_{j4} u_j^{(1)}$
- ...
- $S_j^{(12)} = \varphi_{j0} + \varphi_{j1} S_j^{(11)} + \varphi_{j2} S_j^{(10)} + u_j^{(12)} + \theta_{j1} u_j^{(11)} + \theta_{j2} u_j^{(10)} + \theta_{j3} u_j^{(9)} + \theta_{j4} u_j^{(8)}$

and plug in the resulting  $S_j^{(t)}$  values in (23).

Case 2 ( $j = 9$ , first differences applied; input for formula (24)): Let us consider times  $t = 1, 2, \dots, T$ , where the time increment will correspond to 1 month and  $T = 12$  for a full year. Pay attention to  $p_9 = 1$  and  $q_9 = 3$ . We will evaluate  $D_j(t) = S_j^{(t)} - S_j^{(t-1)}$  starting with initial value  $D_j^{(0)}$ , and evaluating  $D_j^{(t)}$  for all 12 time steps. If the initial values are not available, in accordance with the simplifying assumption above, take  $D_j^{(0)} = 0$ .

The algorithm is based on the ARIMA(1,0,3) model for differences  $D_j$ ,  $j = 9$ :

$$D_j^{(t)} = \varphi_{j0} + \varphi_{j1}S_9^{(t-1)} + u_j^{(t)} + \sum_{i=1}^3 \theta_{ji}u_j^{(t-i)}. \quad (26)$$

We will implement 12 steps:

- $D_j^{(1)} = \varphi_{j0} + \varphi_{j1}D_j^{(0)} + u_j^{(1)}$
- $D_j^{(2)} = \varphi_{j0} + \varphi_{j1}D_j^{(1)} + u_j^{(2)} + \theta_{j1}u_j^{(1)}$
- $D_j^{(3)} = \varphi_{j0} + \varphi_{j1}D_j^{(2)} + u_j^{(3)} + \theta_{j1}u_j^{(2)} + \theta_{j2}u_j^{(1)}$
- $D_j^{(4)} = \varphi_{j0} + \varphi_{j1}D_j^{(3)} + u_j^{(4)} + \theta_{j1}u_j^{(3)} + \theta_{j2}u_j^{(2)} + \theta_{j3}u_j^{(1)}$
- ...
- $D_j^{(12)} = \varphi_{j0} + \varphi_{j1}D_j^{(11)} + u_j^{(12)} + \theta_{j1}u_j^{(11)} + \theta_{j2}u_j^{(10)} + \theta_{j3}u_j^{(9)}$

and plug in the resulting  $D_j^{(t)}$  values in (24).

**Liabilities.** As follows from the simplified model(11), with the parameter values  $\beta_{j,r}$  specified in advance for all  $l \times 2 = 6$  losses corresponding to product lines  $j = 1, \dots, l = 3$ ;  $r = 1, 2$ , variables  $X_j$  represent percentage or relative year-end losses and  $u_{j,r}^{(t)}$  for  $j = 1, \dots, l = 3$ ;  $r = 1, 2$  and  $t = 1, \dots, 12$  are simulated values from the copula model,

$$C_{j,r} = \sum_{t=1}^{12} u_{j,r}^{(t)} = \beta_{j,r}u_{j,r}^{(0)} + (1 + \beta_{j,r}) \sum_{t=1}^{11} u_{j,r}^{(t)} + (1 - \beta_{j,r})u_{j,r}^{(12)}. \quad (27)$$

Thus the results of the simulation for liability variables are directly translated into the end-of-the-year relative liabilities  $X_j$  using the expected claim amount  $E[C_{j,r}]$  and reserve amount  $L_j^{(0)}$ :

$$X_{j,r} = \frac{C_{j,r} - E[C_{j,r}]}{L_j^{(0)}}.$$

Finally, when all  $X_{j,r}$  are evaluated according to (2), the end-of-the-year gain/loss on the portfolio value can be calculated as a percentage of reserve

$$Y = \frac{1}{L^{(0)}}(BOF^{(0)} - BOF^{(12)}) = \frac{A^{(0)}}{L^{(0)}} \sum_{j=1}^9 w_j R_j - \sum_{j=1}^3 v_j \sum_{r=1}^2 X_{j,r}, \quad (28)$$

where weights  $w_j$  can be determined by the investment mix for  $j = 1, \dots, k = 9$  and weights  $v_j$  by the product mix for  $j = 1, \dots, l = 3$ .

## 5 Estimation of Economic Capital

### 5.1 Computer Implementation

Simulation is carried out in R, version 3.5.1, using R-studio, version 1.1.463 (optional). Inputs and outputs are written in Microsoft Excel using VBA macros. Inputs used in the simulation are summarized in *EC Inputs and Results.xlsm* and organized in separate worksheets:

- *Instructions* and troubleshooting
- *EC Results* (defining the confidence level and formulas used for VaR and TVaR and performing EC calculations using updated output of the simulation);
- *Asset/Liability Results* providing full results on relative one-year changes in all asset and liability variables in order of severity using updated output of the simulation, provided for diagnostic purposes;
- *Liability inputs* (including moving average parameter  $\beta$ );
- *Weights* of asset classes and product mix;
- *Correlation table* and copula parameters;
- *Number of years* (scenarios)  $N$ ;
- *Asset parameters* (skewed  $t$  distribution);
- *Initial ARIMA* values;
- *ARIMA parameters* for asset variables;

The executable code in R is attached as *EC Simulation.R*. The simulation code retrieves the required input directly from *EC Inputs and Results.xlsm*. Its output is automatically exported to Microsoft Excel using functions defined in [38] and [39], written in *EC Outputs.xlsx* and exported to the worksheet *EC Results* of the file *EC Inputs and Results.xlsm*.



File *EC Outputs.xlsx* plays a supplementary role and does not have to be analyzed separately other than for monitoring of the model performance. Its column A contains the end-of-the-year values of total gains  $Y$  defined in (28) as a percentage of reserve, sorted in the increasing order. Thus the first several rows of the worksheet represent the high-loss years further used in VaR and TVaR calculation. Each row of the worksheet, along with the total gains in column A, contains components of vector  $(\mathbf{R}, \mathbf{X})$  as defined in (5). These components are listed in columns B-P. Column A is exported to *EC Results* worksheet of *EC Inputs and Results.xlsm* and used in further EC calculations. Columns B-P are exported to worksheet *Asset Liability Results of EC Inputs and Results.xlsm* for diagnostic purposes to determine individual contributions of asset and liability variables to the final EC values.

## 5.2 Example of EC Calculation

In the numerical example considered below, we take  $A_0 = 1.1L_0$  to reflect the initial surplus of assets over reserve. Weights are allocated according to expert suggestions to represent a typical asset allocation and product mix:

$$(\mathbf{w}, \mathbf{v}) = (0, 0, .08, .05, .08, .56, .16, .07, 0, .4, .4, .3, .3, .3, .3). \quad (29)$$

Other inputs reflect the results of parametric estimation and additional assumptions made in Section 3. With the number of years  $N$  set at 1000, we obtain 1000 portfolio gain/loss values from which we can estimate the distribution of the one-year relative gain/loss using Monte Carlo methods. Applying VaR or TVaR to this distribution, we obtain an estimate of EC as the percentage of reserve. An estimate of VaR is given by the fifth worst-case scenario of the total loss and an estimate of TVaR is obtained by averaging the five worst-case scenarios.

Let us summarize the results of multiple runs of simulation of 1000 years and compare them for different values of  $\beta$  in Table 11 and Table 12. For simplicity,  $\beta_{j,r} \equiv \beta$  for all liability variables in one model. Results in Table 11 are based on the current inputs in *EC Inputs and Results.xlsm*, changing  $\beta$  only. Table 12 contains the results of similar calculations after applying the additional 1.8 factor to the standard deviation of total annual claims in the worksheet *Liability Inputs* of the input file. Results of 10 runs with  $N = 1000$  in each are presented in tables as averages and estimated standard deviations. As we see, a moderately conservative assumption in Table 12 does not bring about substantial increase in the EC values without a change in asset assumption. Also, the sensitivity of EC results to the values of parameter  $\beta$  is limited.

Further analysis of high-loss years in *Asset/Liability Results of EC Inputs and Results.xlsm* demonstrates that in some cases high losses occur due to high mortality and morbidity, while in most cases they are driven by big dives in such investment categories as CMBS and junk bonds. In some cases it is a combination of two factors. This gives an argument for the use of copula models directly addressing interdependence of loss factors, especially in case of extreme events.

Table 11: Economic capital calculation (1000 years).

MA parameter	TVaR (st.dev.)	VaR (st.dev.)
$\beta = .1$	-.100 (.020)	-.076 (.013)
$\beta = .5$	-.101 (.027)	-.070 (.011)
$\beta = .99$	-.098 (.016)	-.072 (.009)

Table 12: Economic capital calculation with 1.8 loading(1000 years).

MA parameter	TVaR	VaR
$\beta = .1$	-.099 (.013)	-.080 (.010)
$\beta = .5$	-.102 (.017)	-.079 (.009)
$\beta = .99$	-.109 (.016)	-.082 (.010)

### 5.3 Possible Modifications of the Model Inputs

The simulation from the previous section, resulting in the EC values in Table 11 and Table 12, allows for customization at several different levels. The provided software accepts changes to the inputs, which can reflect the individual experience of the model user without serious adjustment of the code. The following changes can be entered directly in the input file *EC Inputs and Results.xlsm*:

- The number of simulated years,  $N$ , can be chosen and entered in worksheet *Number of Years*. With current settings, it takes less than 2 minutes on a standard desktop computer to run 1000 simulations, and about 10 minutes to run 10,000;
- The confidence level for VaR and TVaR is set at 99.5% to reflect 1 in 200 years events, but can be adjusted in worksheet *EC Results*;
- Weights  $w_j$  and  $v_j$  can be adjusted to reflect the user's asset portfolio allocation and the product mix (worksheet *Weights*);
- Annual values of  $E(C_{j,r})$ ,  $Var(C_{j,r})$ , and  $L_j^{(0)}$  can be chosen to reflect the individual liability profile of the end user in terms of annual expected loss and total reserve amount for each line of business and cause of the loss (mortality, morbidity, lapse). See worksheet *Liability Inputs*.
- Parameter  $\beta$  (worksheet *Liability Inputs*) can be adjusted in the interval  $0 \leq \beta < 1$  to reflect the experience of the end user.  $\beta = 0$  corresponds to month-to-month independence, while the higher values of  $\beta$  represent higher month-to-month autocorrelation.

One advantage of the model setting is the possibility to utilize monthly liability data if they become available. These data may be used to estimate parameters  $E(X_{j,r})$ ,  $Var(X_{j,r})$  and  $\beta_{j,r}$  for each  $j, r$ , and these estimates can serve as model inputs. If the

credibility of these monthly loss data is limited, a Bayesian approach can be recommended to make use of these data along with the initial model parameters, applied in the previous subsection.

Finally, there exist further customization options. Each of these options requires the user to make some changes to the code applied in the model simulation.

- Using marginal distributions for losses other than log-normal (for instance, composite log-normal/Pareto models).
- Using copulas other than Student's  $t$ -copula (e.g., hierarchical Archimedean copulas).
- Adding new classes to the asset portfolio or new products to the mix.
- If necessary, asset and liability variables can be updated on a daily basis, which may be important when considering universal life products or variable annuities.

#### **5.4 Model Assumptions and Simplifications**

The following assumptions and simplifications have to be clarified in order to better understand the scope and limitations of the suggested approach.

Assets:

- Barclay indices are used to represent a typical investment mix of an insurance company.
- The last 15 years of data are used for the estimation of asset and copula parameters to represent a low-interest environment.
- The monthly averages of daily settlement prices are used to estimate asset parameters and autocorrelations.

Liabilities:

- Three simple, traditional products are used to represent a small company's offerings.
- The products used are not overly sensitive to interest rates or asset values.
- The choice of Solvency II parameters as liability distribution drivers is somewhat arbitrary.
- The values used for the correlation of assets and liabilities in matrix form is not verifiable.
- Liability losses assume a degree of autocorrelation that was best estimate.

- Modeled reserves are treated for simplicity as 'stationary state' across time. These factors should allow the assumption that reserve changes will be relatively small compared to the model's loss drivers. As a result, the change in reserve, which is typically an income item, has been ignored for our purposes. This approach seems logical for traditional products, such as the sample products we have used to illustrate the model framework.
- Calibration of liability distributions is based on a 1 in 200 year event severity.
- Reserve and claim relationships are based on simple product design and age grouping.
- Experience and valuation assumptions are equivalent/simplified in some product cases.

## 5.5 Model Validation

The platforms used, R Studio and Microsoft Excel, are widely used and assumed to be both appropriate and, to a high degree, free of errors. No challenger or parallel models were employed for the purpose of validating the software.

Input:

- Mortality and morbidity tables are taken from <http://www.soa.org> and assumed valid;
- Solvency II parameters are widely accepted in the U.S. for baseline guidance by actuaries;
- Industry conventional wisdom and best practice are relied upon to perform reasonableness checks;
- Inspection of intermediate results by an actuary helps to provide a gut check on ranges of inputs and intermediate result output, such as commutation functions used to calculate reserves;
- Benchmarking of values to industry values is performed, such as reserves/1000 for the liability values in the model.

Output:

- EC values were compared to standard benchmarks, conventional wisdom and ranges of results seen in practice;
- Sensitivity testing of EC results to changes in key liability assumption drivers (such as level of assumed mortality and strength of autocorrelation parameters) was performed and an inspection of the results demonstrated reasonable cause and effect relationships.

Process:

- Each sub-process was run separately and checked for reasonableness;
- Errors resulting from initial failed attempts at using model processes and sub-processes were understood and corrected;
- R code was compared to other public domain code for best practice and minimization of error;
- The simulation was run multiple times for N=1000 years per calculation on different computers, with results recorded in Table 11 and Table 12 demonstrating numerical stability, and also for N=10,000 years for additional diagnostic purposes;
- The simulation for N=1000 takes about 100 seconds on a typical laptop computer, thus this sample size can be recommended;
- The developers did not attempt replication of model results using challenger models and relied, instead, upon reasonableness of sub-process results and aggregate process outputs produced;
- Ongoing monitoring is the validation step taken each time a model is updated and rerun and should include some of the key validation steps performed during the development stage. This is often performed by a person other than the developer. The developers of this model assume that the end user takes on this responsibility.

Further model validation can be carried out in four ways:

- Sensitivity tests using admissible changes of liability inputs (reflecting more and less conservative assumptions);
- Sensitivity tests using admissible changes in asset allocation and product mix;
- Analysis of the asset/liability outcomes for the contribution of certain asset and liability classes to the critical values of overall loss in the extreme scenarios provided in *EC Inputs and Results.xlsm*;
- Running a larger number of scenarios (years) (100,000 or 1,000,000).

In all four cases we are looking for a relative stability of the tail values and their sensitivity to the changes of model inputs.

## 6 Acknowledgements

The work is supported by the research grant program of the Society of Actuaries (SOA). The authors are grateful to Heekyung Youn for her valuable comments and insights, our industry consultants Vladimir Ladyzhets and David Walczak for providing expert help on variable selection and data research, and also to Thomas Vlasak for his help in asset data analysis. The authors also express sincere gratitude to the Project Oversight Committee for many valuable comments and suggestions.

## References

- [1] Ane, T., Ureche-Rangau, L. and Labidi-Makni, C. Time-varying conditional dependence in Chinese stock markets. *Applied Financial Economics*, 18, 895-916, 2008.
- [2] Azzalini, A. and Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc., series B*, 65, 367-389, 2003.
- [3] Fixed Income Barclay Bloomberg Indices <https://indices/barclays/>.
- [4] Beaulieu, N. C., Abu-Dayya, A. A., and McLane, P. J. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Transactions on Communications*, 43, 2869, 1995.
- [5] Braun, A., Schmeiser, H. and Schreiber, F. Portfolio optimization under solvency II: implicit constraints imposed by the market risk standard formula. *Journal of Risk and Insurance*, 84, 177207, 2017.
- [6] Cardieri, P., and Rappaport, T. S. Statistics of the sum of lognormal variables in wireless communications. *IEEE Vehicular Technology Conference Proceedings*, 3, 1823-1827, 2000.
- [7] Christiansen, M. C. and Niemeyer, A. Fundamental definition of the solvency capital requirement in Solvency II. *ASTIN Bulletin*, 44, 501533, 2014.
- [8] Cooray, K. and Ananda, M. M. Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal*, 5, 321334, 2005.
- [9] Cummins, J. D. and Phillips, R. D. Capital adequacy and insurance risk-based capital systems. *Journal of Insurance Regulation*, 28, 25-72, 2009.
- [10] Demarta, S. and McNeil, A. The t-copulas and related copulas. *International Statist. Review*, 73, 111-129. 2005.
- [11] Dutang, C., Goulet V., and Pigeon, M. actuar: An R Package for Actuarial Science. *Journal of Statistical Software*, vol. 25, no. 7, 1-37. URL <http://www.jstatsoft.org/v25/i07>, 2008.
- [12] Eckert, J. and Gatzert, N. Risk-and value-based management for non-life insurers under solvency constraints. Working Paper, 2017.
- [13] Farr, I. , Mueller, H., Scanlon, M., and Stronkhorst, S. Economic Capital for Life Insurance Companies, *SOA and Towers Perrin*, 2008.
- [14] Frees E. W., and Valdes E. A. Understanding Relationship using Copulas. *North American Actuarial Journal*, 2, 1-25, 1998.
- [15] Galanos, A. Package Rugarch - CRAN-R - R Project. <https://cran.r-project.org/web/packages/rugarch/rugarch.pdf>, 2017.

- [16] Genest C. and Rivest L. P. Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, 88, 423, 1034-1043, 1993.
- [17] Gordeev, V., Knyazev, A., and Shemyakin, A. Selection of Copula Model for Inter-market Dependence. *Model Assisted Statistics and Applications*, 7:4, 315–325, 2012.
- [18] Hansen, B. E. Autoregressive conditional density estimation. *International Economics Review*, 35:3, 705–730, 1994.
- [19] Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. —newblock Copula: Multivariate Dependence with Copulas. R package version 0.999-18, URL <https://CRAN.R-project.org/package=copula>, 2017.
- [20] Hofert, M. and Maechler, M. Nested Archimedean copulas meet R. *Journal of Statistical Software*, 39:9, 1–20, 2011.
- [21] Kangina, N., Kniyazev, A., Lepikhin, O., and Shemyakin, A. Modeling joint distribution of national stock indices. *Model Assisted Statistics and Applications*, 11:1, 2016.
- [22] Klein, M. Long-Term Rate Outlook Is Surprisingly Tame. Barrons, June 2018. <https://www.barrons.com/articles/long-term-rate-outlook-is-surprisingly-tame-1529098336>
- [23] Kniyazev, A., Lepikhin, O., and Shemyakin, A. Joint distribution of stock indices: methodological aspects of construction and selection of copula models. *Applied Econometrics*, 42, 1, 30-53, 2016.
- [24] Low Interest Rates, The Center for Insurance Policy and Research, National Association of Insurance Commissioners, March 2018. <https://www.naic.org/cipr-topics/topic-low-interest-rates.htm>
- [25] McKenzie, D. and Spears, T. The formula that killed Wall Street: The Gaussian copula and modelling practices in investment banking. *Social Studies of Science*, 44, 393-417, 2014.
- [26] Morone M., Cornaglia A., and Mignola G. Economic Capital Assessment via Copulas: Aggregation and Allocation of Different Risk Types. [http://www.defaultrisk.com/pp\\_model160.htm](http://www.defaultrisk.com/pp_model160.htm), 2009.
- [27] Nadarajah, S., and Bakar, S.A.A. New Composite Models for the Danish Fire Insurance Data. *Scandinavian Actuarial Journal*, 2, 180-187, 2014.
- [28] Nadarajah, S. CompLognormal: Functions for actuarial scientists. R package version 3.0. <https://CRAN.R-project.org/package=CompLognormal>, 2013.
- [29] Nguyen T. and Molinari R. D. Risk Aggregation by using Copulas in Internal Models. *Journal of Mathematical Finance*, 1, 50-57, 2011.

- [30] Okhrin, O., and Ristig, A. Hierarchical Archimedean Copulae: The HAC Package, *Journal of Statistical Software*, 58:4), 1-20. URL <http://www.jstatsoft.org/v58/i04/>, 2014.
- [31] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>, 2018.
- [32] Schepsmeier, U. Stoeber, J., Brechmann, E. C., Graeler, B., Nagler T., and Erhardt, T. VineCopula: Statistical Inference of Vine Copulas. R package version 2.1.6. <https://CRAN.R-project.org/package=VineCopula>, 2018.
- [33] Schwartz, S. C., and Yeh, Y. S. On the distribution function and moments of power sums with lognormal components. *Bell System Technical Journal*, 61, 1441-1462, 1982.
- [34] Scollnik, D.P. and Sun, C. Modeling with Weibull-Pareto models. *North American Actuarial Journal*, 16, 260272, 2012.
- [35] Shemyakin, A., and Kniazev, A. *Introduction to Bayesian Estimation and Copula Models of Independence*. Wiley, London/New York, 2017.
- [36] Shim J., Lee S., and MacMinn R. Measuring Economic Capital: Value-at-Risk, Expected Shortfall and Copula Approach. <http://dx.doi.org/10.2139/ssrn.1840124>, 2011.
- [37] U.S. Insurance Industry Cash and Invested Assets at Year 2016. [http://www.naic.org/capital\\_markets\\_archive/170824.htm](http://www.naic.org/capital_markets_archive/170824.htm), 2017.
- [38] Walker, A. openxlsx: Read, Write and Edit XLSX Files. R package version 4.1.0. <https://CRAN.R-project.org/package=openxlsx>, 2018
- [39] Wickham, M. and Bryan, J. readxl: Read Excel Files. R package version 1.1.0. <https://CRAN.R-project.org/package=readxl>, 2018.