Modeling Multivariate Risk

To Copula, or Not To Copula: That is the Question

X. Sheldon Lin

Department of Statistics
University of Toronto
Email: sheldon@utstat.utoronto.ca

Joint work with Simon Lee

46th Actuarial Research Conference
University of Connecticut
August 11-13, 2011

Outline

Motivation

A good multivariate model will:

- identify the level of dependence between insurance portfolios/blocks.
- provide accurate assessment of the risk exposure of an insurance portfolio.
- help examine the diversification effect among and within the portfolios.
- determine required capitals/reserves (regulatory and internal) using appropriate risk measures.
- be useful in solvency/capital adequacy tests.

Copula Methodology

- The most popular methodology in multivariate modeling in finance and insurance.
- Extremely easy to understand.
- Another advantage of the copula approach is that it uses a two stage procedure that separates the dependence structure of a model distribution from its marginals.

What is a copula?

An k-dimensional copula $C(\mathbf{u})$ with $\mathbf{u} = (u_1, \cdots, u_k)$ is a real-valued function defined on the n-dimensional unit cube $\mathbb{I}^k$, where $\mathbb{I} = [0, 1]$, that has the following properties:

- $C(\mathbf{u}) = 0$, if at least one of the coordinates is 0;
- $C(\mathbf{u}) = u_k$, if all other coordinates are 1;
- For any n-dimensional box $[\mathbf{a}, \mathbf{b}]$, where $\mathbf{a} = (a_1, \cdots, a_k)$ and $\mathbf{b} = (b_1, \cdots, b_k)$, the volume

$$\Delta_{a_k}^{b_k} \cdots \Delta_{a_1}^{b_1} C(\mathbf{u}) > 0.$$

In other words, $C(\mathbf{u})$ is a joint distributional function with uniform marginals.

Sklar's Theorem

For any joint distribution function $F(\mathbf{x})$ with marginals $F_1(x_1), \cdots, F_k(x_k)$, there exists a k-dimensional copula $C(\mathbf{u})$ such that

$$F(\mathbf{x}) = C\left(F_1(x_1), \cdots, F_k(x_k)\right).$$

The Use of Copulas

- The two stage procedure: one to estimate the marginals and the other to choose a copula to determine the dependent structure, according to Sklar's Theorem.
- Key is to construct a copula that can capture the dependent structure of a given dataset.
- Many choices for a two-dimensional copula:
  Archimedean copulas: Clayton, Ali-Mikhail-Haq, Gumble, Frank; Farlie-Gumble-Morgensten; Gaussian; Empirical; ....
- Few higher-dimensional copulas are available.
- An excellent reference:
  E.W. Frees and E.A. Valdez (1998). "Understanding relationships using copulas", *North American Actuarial Journal*, 2(1), 1-25.
- Question: Is the copula methodology always desirable for modeling dependency?

Some Properties of an Ideal Multivariate Model

Quotes from Joe H. (1997). *Multivariate Models and Dependence Concepts,* Chapman and Hall, London:
An ideal multivariate parametric model should have the following four desirable properties

A. interpretability, which could mean something like mixture, stochastic or latent variable representation;
B. the closure property under the taking of margins, in particular the bivariate margins belonging to the same parametric family (this is especially important if, in statistical modeling, one thinks first about appropriate univariate margins, then bivariate and sequentially to higher-order margins);

Some Properties of an Ideal Multivariate Model

C. a flexible and wide range of dependence (with type of dependence structure depending on applications);

D. a closed-form representation of the cdf and density (a closed-from cdf is useful if the data are discrete and a continuous random vector is used), and if not closed-form, then a cdf and density that are computationally feasible to work with.

How about a Copula Model?

- Property C is often not satisfied for most copulas. This is because the dependence structure is predetermined in a copula. Fitting to data with complicated features such as multiple modes could be unsatisfactory.
- Property D is not easily satisfied either. In many cases, the cdf and some other quantities of interest of a multivariate distribution based on a copula may not be obtained explicitly. As a result, simulation is often the only tool available.
- Dimensionality is another potential problem. Although this is not unique to copulas, it seems that copulas make the problem worse in general. This might be the reason that the dominating majority of copula applications so far are limited to bivariate cases. However, in insurance we often need to model dependence among a large number of correlated business blocks, which can be difficult to tackle by a copula method.
- Some criticisms can be found in Mikosch, T. (2006). "Copulas: tales and facts," *Extremes*, 9, 3-20.

An Alternative

- Model the dependence directly using a multivariate parametric model

Proposed Model: Multivariate Erlang Mixture

The density of a k-variate Erlang mixture is of the form:

$$f(\mathbf{x}|\theta, \boldsymbol{\alpha}) = \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} p(x_j; m_j, \theta),$$

where

$$p(x; m, \theta) = \frac{x^{m-1} e^{-x/\theta}}{\theta^m (m-1)!},$$

$\mathbf{x} = (x_1, \cdots, x_k)$, $\mathbf{m} = (m_1, \cdots, m_k)$,
$\boldsymbol{\alpha} = (\alpha_{\mathbf{m}}; m_i = 1, 2, \cdots; i = 1, 2, \cdots, k)$ with each $\alpha_{\mathbf{m}} \geq 0$ and
$\sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} = 1$.

Could the Erlang Mixture be a good Multivariate Model?

- It is a natural extension of the univariate Erlang mixture but is it a good model?
- The class of multivariate Erlang mixtures is dense in the space of positive continuous multivariate distributions.
- In theory we can fit a multivariate Erlang mixture to any multivariate data within a given accuracy.

Expectation-Maximization (EM) Algorithm

A MLE based algorithm for incomplete data.

- Let $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be an incomplete sample generated from a pair of random variables/vectors (X, Y) with joint density $p(x, y|\Phi)$, where Y is an unobservable random variable and Φ is the set of parameters to be estimated.
- The complete-data log-likelihood is given by

$$l(\Phi|\mathbf{x}, \mathbf{Y}) = \sum_{i=1}^{n} \ln p(x_i, Y_i|\Phi)$$

- Given the sample $\mathbf{x}$ and the current estimate of the parameters $\Phi^{(k-1)}$, the posterior distribution of Y_i is given by

$$q(y_i|x_i, \Phi^{(k-1)}) = \frac{p(x_i, y_i|\Phi^{(k-1)})}{p(x_i|\Phi^{(k-1)})},$$

where $p(x|\Phi^{(k-1)})$ is the marginal density.

Expectation-Maximization (EM) Algorithm

- The expected posterior log-likelihood (E-Step) is given by

$$
\begin{aligned}
Q(\Phi|\Phi^{(k-1)}) &= \sum_{i=1}^{n} E\{\ln p(x_i, Y_i|\Phi)\} \\
&= \sum_{i=1}^{n} \int [\ln p(x_i, y_i|\Phi)] q(y_i|x_i, \Phi^{(k-1)}) dy_i
\end{aligned}
$$

- Maximize the log-likelihood (M-Step):

$$
\Phi^{(k)} = \max_{\Phi} Q(\Phi|\Phi^{(k-1)})
$$

An EM Algorithm for Finite Erlang Mixtures

- Data fitting is easy as an EM algorithm is available.
- Data set of k dimensions: $\mathbf{x}_v = (x_{1v}, x_{2v}, \cdots, x_{kv}), \ v = 1, \cdots, n$. We are to use a k-variate finite Erlang mixture to fit the data.
- Parameters to be estimated (denoted by Φ): the scale parameter θ and all the mixing weights $\alpha_{\mathbf{m}}$'s, where the shape parameters $\mathbf{m}$'s are initially preset and denoted by $\mathcal{M}$. If $\mathbf{m} \notin \mathcal{M}$, we set $\alpha_{\mathbf{m}} = 0$.

The EM Algorithm

For $\mathbf{m} \in \mathcal{M}$,

$$q(\mathbf{m}|\mathbf{x}_v, \Phi^{(l-1)}) = \frac{\alpha_{\mathbf{m}}^{(l-1)} \prod_{j=1}^{k} p(x_{jv}, m_j, \theta)}{\sum_{r_1=1}^{\infty} \cdots \sum_{r_k=1}^{\infty} \alpha_{\mathbf{r}}^{(l-1)} \prod_{j=1}^{k} p(x_{jv}, r_j, \theta)}$$

$$\alpha_{\mathbf{m}}^{(l)} = \frac{1}{n} \sum_{v=1}^{n} q(\mathbf{m}|\mathbf{x}_v, \Phi^{(l-1)}), \ \mathbf{m} \in \mathcal{M},$$

and

$$\theta^{(l)} = \frac{\sum_{v=1}^{n} \sum_{j=1}^{k} x_{jv}}{n \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \left(\sum_{j=1}^{k} m_j \right) \alpha_{\mathbf{m}}^{(l)}}$$

The EM Algorithm: Initial Estimation and Shape Parameter Adjustment

- Us an “80-8” rule to choose an initial value of θ. After the value of θ is set, the empirical distribution is used to determine the value of each $\alpha_{\mathbf{m}}$.
- Run the EM algorithm to initially fit the data and reduce the number of components in the mixture.
- Adjust the shape parameters by increasing or decreasing their values and run the EM algorithm repeatedly. Use Schwarz’s Bayesian Information Criterion (BIC) to further reduce the number of components in the mixture.

A Preliminary Numerical Experiment

- Fitting data generated from a multivariate log normal distribution of 12 dimensions.
- Let

$$X_i = \prod_{j=1}^{i} Z_j, \; i = 1, 2, \cdots, 12,$$

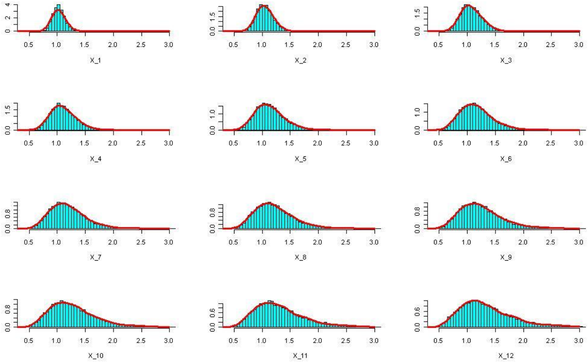
  where $Z_j, j = 1, 2, \cdots, 12$, be iid log normal random variables with parameters μ and σ.
  $(X_1, \cdots, X_{12})$ has a multivariate log normal distribution.
- This example is motivated by the applications in the pricing of arithmetic Asian options and equity-indexed annuities (EIA). Consider the price of a risky asset or an equity index that follows a geometric Brownian motion with drift 12μ and volatility 12σ over a one-year period. Thus, $X_1, \cdots, X_{12}$ represent the prices of the asset at the end of each month.
- Assume that $\mu = 2.5\%$ and $\sigma = 10\%$ and simulate 8000 observations from $(X_1, X_2, \cdots, X_{12})$.

Parameter values

	m_{i_1}	m_{i_2}	m_{i_3}	m_{i_4}	m_{i_5}	m_{i_6}	m_{i_7}	m_{i_8}	m_{i_9}	$m_{i_{10}}$	$m_{i_{11}}$	$m_{i_{12}}$	$\alpha_{\mathbf{m}}$
1	75	70	65	62	59	57	55	54	53	52	52	52	0.03519954
2	77	75	73	72	73	75	78	82	86	91	97	101	0.06750167
3	75	70	66	64	63	63	64	65	68	70	73	75	0.05352882
4	80	79	79	81	83	86	91	98	106	115	122	129	0.06488830
5	80	81	84	89	96	103	109	113	114	114	112	112	0.06019880
6	83	86	90	94	99	105	111	120	129	138	145	150	0.08021910
7	80	79	78	77	75	72	69	66	64	62	61	61	0.06330692
8	79	78	77	77	77	77	77	77	77	78	79	80	0.11508296
9	82	83	84	86	87	89	91	92	94	94	95	97	0.13055435
10	85	88	94	100	109	119	129	143	158	171	182	191	0.03218294
11	89	99	109	116	125	133	139	143	146	149	152	156	0.04549171
12	85	89	92	93	92	90	87	83	79	77	76	76	0.05818215
13	87	92	97	99	100	100	100	102	105	110	116	121	0.06408133
14	87	93	99	103	105	106	105	102	99	96	93	93	0.05392744
15	88	96	104	112	119	122	123	123	122	122	121	122	0.05431533
16	91	103	114	128	141	156	167	178	189	199	209	214	0.02133865

Table: The shape parameters and estimated weights of the fitted distribution with $\theta = 0.01253039$

Fitting Marginals



Aggregated Loss

- The validity of using the marginals to represent the fitness of the model is questionable as the dependence structure is not shown in these plots.
- To address the issue, we investigate the fitness of the density of $S_{12} = X_1 + X_2 + \cdots + X_{12}$ that is a univariate Erlang mixture as shown later on.
- Since a poor overall fitting to the multivariate data would in general result in a poor fitting to the aggregated data, fitting to the aggregated data could be a good measure for the goodness of fit. The next 3 slides provide the fitting results in this regard.

Histogram of Aggregated Data

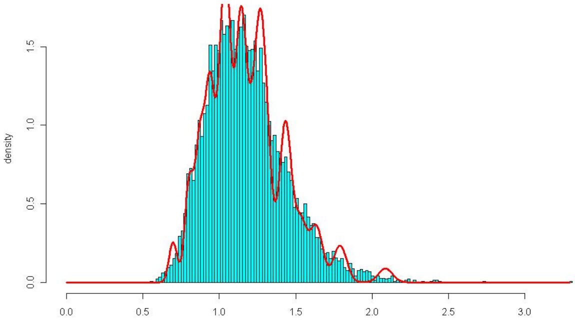


Figure: Histogram of the aggregated data and the density of the fitted distribution

Goodness-of-Fit Tests

Test	Statistic	p-value	Accepted at 5% significant level?
Chi Square Test	818.32	0.3099	Yes
K-S Test	0.05	0.27	Yes
AD Test	0.4378	0.2228	Yes

Comparison of Moments

Moment	Empirical Distribution	Fitted Distribution	Fitted/ Empirical	Percentage Difference (%)
1	1.1791	1.1791	1.00000	0.0000%
2	1.4566	1.4588	0.9985	0.1511%
3	1.8871	1.8971	0.9947	0.5284%
4	2.5654	2.5985	0.9829	1.2712%
5	3.6605	3.7592	0.9737	2.6237%

Table: The first 5 moments of the empirical and fitted distributions

Distributional Properties

- Let the random vector $\mathbf{X} = (X_1, \cdots, X_k)$ follow a multivariate Erlang mixture and $\mathbf{N} = (N_1, \cdots, N_k)$ be a multivariate counting random vector with probability function

$$P(\mathbf{N} = \mathbf{m}) = \alpha_{\mathbf{m}}, \;\; m_j = 1, 2, \cdots; \;\; j = 1, \cdots, k.$$

  Then, the characteristic function of $\mathbf{X}$ is given by

$$\varphi(\mathbf{z}) \quad = \quad P_{\mathbf{N}}\left(\frac{1}{1 - \mathbf{i}\theta z_1}, \cdots, \frac{1}{1 - \mathbf{i}\theta z_k}\right).$$

  where $P_{\mathbf{N}}(\mathbf{z})$ is the probability generating function of $\mathbf{N}$.
- A multivariate Erlang mixture is a multivariate compound exponential distribution.
- The marginal distribution of X_j is a univariate Erlang mixture. The weights of the mixture are

$$\alpha_{m_j}^{(j)} \stackrel{\text{def}}{=} \sum_{m_l, l \neq j} \alpha_{\mathbf{m}}.$$

  Furthermore, any p-variate $(p < k)$ marginal is a p-variate Erlang mixture.

Distributional Properties

- The marginal random variables $X_1, \cdots, X_k$ are mutually independent if the counting random variables $N_1, \cdots, N_k$ are mutually independent. In this case, we have

$$\alpha_{\mathbf{m}} = \prod_{j=1}^{k} \alpha_{m_j}^{(j)},$$

where $\{\alpha_{m_j}^{(j)},\ m_j = 1, 2, \cdots,\}$ is the distribution of N_j.
- The sum $S_k = X_1 + \cdots + X_k$ has a univariate Erlang mixture with the mixing weights being the coefficients of the power series $P_{\mathbf{N}}(z, \cdots, z)$: for $i = 1, 2, \cdots,$

$$\alpha_i^S = \sum_{m_1 + \cdots + m_k = i} \alpha_{\mathbf{m}}.$$

Multivariate Excess Losses

Let $\mathbf{d} = (d_1, \cdots, d_k)$ be deductible levels (or economic capitals) of the individual losses $\mathbf{X} = (X_1, \cdots, X_k)$ from an insurance portfolio.

The associated multivariate excess losses may thus be defined as the conditional random vector $\mathbf{Y_d} = \mathbf{X} - \mathbf{d} | \mathbf{X} > \mathbf{d}$.

The joint density of $\mathbf{Y_d}$ is again a multivariate Erlang mixture with the same scale parameter. Its mixing weights are given by:

$$c_{\mathbf{i}} = \frac{\theta^k}{\overline{F}(\mathbf{X} > \mathbf{d})} \sum_{m_1=i_1}^{\infty} \cdots \sum_{m_k=i_k}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} p(d_j; m_j - i_j + 1, \theta).$$

Multivariate Excess Losses

- This result allows for explicit calculation of VaR and TVaR of individual losses simultaneously!
- If X_i is interpreted as the time of default of Firm i and $d_1 = \cdots = d_k = t$, then the distribution is the joint distribution of the default times, given that all firms survive to time t.

Moment Properties

- The joint moment

$$E\left\{\prod_{j=1}^{k} X_j^{n_j}\right\} = \theta^n \sum_{m_1=1}^{\infty} \cdots \sum_{m_k=1}^{\infty} \alpha_{\mathbf{m}} \prod_{j=1}^{k} \frac{(m_j + n_j - 1)!}{(m_j - 1)!},$$

  where $n = \sum_{j=1}^{k} n_j$.
- **Covariance Invariance** The covariance of any marginal pair (X_j, X_l) is proportional to the covariance of (N_j, N_l). More precisely,

$$Cov(X_j, X_l) = \theta^2 Cov(N_j, N_l).$$

Dependence Measure: Kendall's tau

- Kendall's tau for a pair of continuous random variables X and Y measures the tendency that X and Y will move in the same direction (concordance).
  It is defined as

  $$\tau = P\left\{(X_1 - X_2)(Y_1 - Y_2) > 0\right\} - P\left\{(X_1 - X_2)(Y_1 - Y_2) < 0\right\},$$

  where (X_1, Y_1) and (X_2, Y_2) are two iid copies of (X, Y).
- Unlike the (Pearson) correlation coefficient, it does not assume linear relationship. In this regard, Kendall's tau is more meaningful in measuring the correlation between two random variables.

Dependence Measure: Kendall's tau

Kendall's tau of a bivariate Erlang mixture is given by

$$\tau = 4 \sum_{i,j=0}^{\infty} \sum_{k,l=1}^{\infty} \binom{i+k-1}{i} \binom{j+l-1}{j} \frac{Q_{ij}\alpha_{kl}}{2^{i+j+k+l}} - 1,$$

where $Q_{ij} = \sum_{k=i+1}^{\infty} \sum_{l=j+1}^{\infty} \alpha_{kl}$ is the survival function of the mixing distribution.

Dependence Measure: Spearman's rho

- Spearman's rank correlation coefficient (Spearman's rho) is another commonly used measure of association.
  It is defined as

  $$\rho = 3(P\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_3) < 0\}),$$

  where (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) are iid copies of (X, Y).
- Spearman's rho of a bivariate Erlang mixture is given by

  $$\rho = 12 \sum_{i,j=0}^{\infty} \sum_{k,l=1}^{\infty} \binom{i+k-1}{i} \binom{j+l-1}{j} \frac{Q_{ij}\alpha_k^{(1)}\alpha_l^{(2)}}{2^{i+j+k+l}} - 3$$

Aggregate Losses

- The sum $S_k = X_1 + \cdots + X_k$ has a univariate Erlang mixture with the mixing weights being
$$\alpha_i^S = \sum_{m_1+\cdots+m_k=i} \alpha_{\mathbf{m}}.$$
- The value-at-risk at confidence level p, $V = VaR_p(S_k)$, is the solution of equation
$$e^{-V/\theta} \sum_{i=0}^{\infty} Q_i \frac{V^i}{\theta^i i!} = 1 - p, \quad Q_i = \sum_{j=i+1}^{\infty} \alpha_j^S.$$
- The Tail VaR at confidence level p, $TVaR_p(S_k)$, is given by
$$TVaR_p(S_k) = \frac{\theta e^{-V/\theta}}{1-p} \sum_{i=0}^{\infty} Q_i^* \frac{V^i}{\theta^i i!} + V, \quad Q_i^* = \sum_{j=i}^{\infty} Q_j.$$
- The stop-loss premium of S_k at deductible level d, $E\{(S_k - d)_+\}$ is given by
$$E\{(S_k - d)_+\} = \theta e^{-d/\theta} \sum_{i=0}^{\infty} Q_i^* \frac{d^i}{\theta^i i!}.$$

Potential Financial Applications

- Option Pricing: Discrete Lookback, Asian, Basket...
- Default Risk Modeling
  Gaussian models are commonly used to model/fit positive data. Often a highly nonlinear transformation is required if we do so. Example: modeling the default times of firms using a Gaussian copula. Instead of mapping the distribution of a default time to a Gaussian distribution in a non-linearly way, we may use the multivariate model to fit default time data directly.

Questions?

The results in this presentation and more and be found in
Lee, S.C.K. and Lin, X.S. (2011). "Modeling dependent risks with multivariate Erlang mixtures," ASTIN Bulletin, under revision.

Thank you for listening. Your turn now.....