RT-MAE 2007-06

COPULA-BASED REGRESSION MODELS

by

Nikolai Kolev
and
Delhi Paiva

Palavras-Chave: Archimedian copula, Copula theory, Estimation, Frank copula, Gaussian copula, Insurance data, Longitudinal data, Markov processes, Multivariate generalized hyperbolic distributions, Regression models, Semiparametric specifications, Simulated maximum likelihood, Time series analysis, Transition models, Unobserved heterogeneity.
Classificação 60E05, 62H05, 62H10, IE43, IM11, IM12.

- Fevereiro de 2007 -

Copula-Based Regression Models

Nikolai Kolev[1] and Delhi Paiva[2]
[1]Department of Statistics, University of São Paulo
Cx. Postal 66.281, 05311-970 - São Paulo, SP, Brazil
[2]EACH, University of São Paulo
nkolev@ime.usp.br and delhi@usp.br

Abstract

We summarize main facts about copula theory, discussing its pitfalls. Several copula-based regression models are represented. We show the advantages of the trivariate Frank copula model in comparison to the usual regression analysis applied to real insurance data. Multivariate distribution models with generalized hyperbolic marginals and their copulas are discussed along with applications in a longitudinal data framework.

Key-words: Archimedian copula, Copula theory, Estimation, Frank copula, Gaussian copula, Insurance data, Longitudinal data, Markov processes, Multivariate generalized hyperbolic distributions, Regression models, Semiparametric specifications, Simulated maximum likelihood, Time series analysis, Transition models, Unobserved heterogeneity.

MSC: 60E05, 62H05, 62E10, IE43, IM11, IM12.

1 Copulas

The shortest definition (characterization) of the copula function is as follows: a copula is a multivariate distribution whose marginal distributions are uniformly distributed on the interval (0,1). The importance of copulas in modelling the distribution of multivariate random variables is justified by Sklar's theorem. For simplicity, we will consider the bivariate case in this section.

1.1 Continuous case

Let $F(y_1, y_2)$ denote the joint distribution function of continuous random variables Y_1 and Y_2 with marginal distributions $F_1(.)$ and $F_2(.)$ respectively. Denote by $F_i^{-1}(v_i) = inf\{y_i \in \Re, F_i(y_i) \geq v_i\}$ the corresponding inverse functions, $i = 1, 2$. Sklar's theorem, (see e.g., Sklar (1959)), states that there exists a unique copula function

$$C(v_1, v_2) = F(F_1^{-1}(v_1), F_2^{-1}(v_2)), \quad v_1, v_2 \in (0, 1)$$

that connects $F(y_1, y_2)$ to $F_1(y_1)$ and $F_2(y_2)$ via

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2)), \quad y_1, y_2 \in (-\infty, \infty). \tag{1}$$

Hence, the information in the joint distribution $F(y_1, y_2)$ is decomposed into those of marginal distributions and that of copula function $C(.,.)$ which captures the dependence structure between Y_1 and Y_2.

On the other hand, for any copula function $C(.,.)$ and any univariate distribution functions $F_1(y_1)$ and $F_2(y_2)$, the function $C(F_1(y_1), F_2(y_2))$ is a bivariate distribution function $F(y_1, y_2)$ as given by (1).

Consequently, copulas allow one to model the marginal distributions and the dependence structure of multivariate random variable separately. The copula function is therefore a class of multivariate distributions that has the appealing feature of being functionally independent of its marginals (which are uniformly distributed on $(0, 1)$). Such a family of distributions can be expressed as a function of arbitrary marginal distributions.

Example 1. Let Y_1 and Y_2 be random variables with joint distribution given by

$$F(y_1, y_2) = [1 + \exp(-y_1) + \exp(-y_2)]^{-1},$$

for all $y_1, y_2 \in [-\infty, \infty]$, i.e. $F(y_1, y_2)$ is the *Gumbel bivariate logistic distribution*. The marginal distributions are

$$F_1(y_1) = [1 + \exp(-y_1)]^{-1} \quad \text{and} \quad F_2(y_2) = [1 + \exp(-y_2)]^{-1},$$

hence

$$F_1^{-1}(v_1) = -\ln\left(\frac{1 - v_1}{v_1}\right) \quad \text{and} \quad F_2^{-1}(v_2) = -\ln\left(\frac{1 - v_2}{v_2}\right).$$

Then, according to Sklar's theorem

$$C(v_1, v_2) = F(F_1^{-1}(v_1), F_2^{-1}(v_2)) = \frac{v_1 v_2}{v_1 + v_2 - v_1 v_2}, \quad v_1, v_2 \in [0, 1].$$

Example 2. If Y_1 and Y_2 be random variables with joint distribution given by

$$F(y_1, y_2) = \begin{cases} \frac{(y_1+1)[\exp(y_2)-1]}{y_1+2\exp(y_2)-1}, & \text{if } (y_1, y_2) \in [-1, 1] \times [0, \infty); \\ 1 - \exp(-y_2), & \text{if } (y_1, y_2) \in (1, \infty) \times [0, \infty); \\ 0, & \text{elsewhere,} \end{cases}$$

then the marginal distributions

$$F_1(y_1) = \frac{y_1+1}{2},\ y_1 \in [-1,1] \quad \text{and} \quad F_2(y_2) = 1 - \exp(-y_2),\ y_2 \geq 0,$$

are uniformly distributed on $(-1, 1)$ and unit exponentially distributed, respectively. The corresponding inverses are

$$F_1^{-1}(v_1) = 2v_1 - 1 \quad \text{and} \quad F_2^{-1}(v_2) = -\ln(1 - v_2).$$

The associated copula $C(v_1, v_2)$ happens to have the same form as in Example 1.

An attractive characteristic of the copula class is that the elimination of the marginals through the copula helps to model and understand dependence structure effectively, as the dependence has no relationship with the marginal behavior (according to the mathematical copula theory). Note, that the copula function in Examples 1 and 2 are the same, nevertheless different marginal distributions, with completely different support.Theoretically, this fact is not surprising, but it is confusing (for practitioners).

Thus, the functional dependence can be studied without the need of specifying the marginal distributions. This is useful for empirical applications because if the bivariate distribution $F(y_1, y_2)$ is unavailable but the univariate marginals $F_1(y_1)$ and $F_2(y_2)$ are known, then an appropriate choice of copula function $C(v_1, v_2)$ in (1) provides a representation of the unknown bivariate distribution. In addition, copula can be specified to include a vector of parameters θ that captures the degree of association between the univariate marginals.

The copula theory has an increasing evolution during the last decade, motivated by its application in probability theory, statistics, finance, insurance, economics, see for example the recent monographs Cherubini et al. (2004), McNeil et al. (2005), Kolev et al. (2006), Nelsen (2006), Trivedi and Zimmer (2006), and references therein. The importance of copula utilization is based on indicating copula methodology as a tool for dependence analysis of financial institutions in Basel II agreement and Solvency II (for insurance companies).

Remark 1 (Estimation using copulas in regression analysis). In regression context, each marginal distribution can be specified to be conditioned on a vector of covariates. Estimation proceeds by first selecting the appropriate copula $C(.,.;\theta)$ depending on its vector of parameters θ and marginal distributions $F_1(y_1|x_1, \beta_1)$ and $F_2(y_2|x_2, \beta_2)$ where x_1 and x_2 are covariates, and β_1 and β_2 are unknown parameters (x_1 and x_2 need not be different sets of covariates).

Then standard maximum likelihood techniques are applied to the joint distribution

$$F(y_1, y_2|x_1, x_2, \beta_1, \beta_2; \theta) = C(F_1(y_1|x_1, \beta_1)), F_2(y_2|x_2, \beta_2); \theta).$$

For notational convenience we suppress for a moment the conditioning on covariates and parameters. If $F_1(.)$ and $F_2(.)$ are continuous with densities $f_1(.)$ and $f_2(.)$, then the corresponding joint density $f(.,.)$ and the copula density $c(.,.)$ for each observation $i = 1, 2, \ldots, n$ are related through equation

$$f(y_{i1}, y_{i2}) = \frac{\partial^2 F(y_{i1}, y_{i2})}{\partial y_{i1} \partial y_{i2}} = \frac{\partial^2 C(F_1(y_{i1}), F_2(y_{i2}))}{\partial F_1(y_{i1}) \partial F_2(y_{i2})} \times \frac{\partial F_1(y_{i1})}{\partial y_{i1}} \times \frac{\partial F_2(y_{i2})}{\partial y_{i2}},$$

which for $\Theta = (\beta_1, \beta_2, \theta)$ gives

$$f(y_{1i}, y_{2i}, \Theta) = c(F_1(y_{1i}, \beta_1), F_2(y_{2i}, \beta_2), \theta) f_1(y_{1i}, \beta_1) f_2(y_{2i}, \beta_2). \tag{2}$$

Next, we present two important families of bivariate copula models which have a number of desirable properties. For additional list of copula functions we refer to Joe (1997).

Gaussian copula. Let $\Phi_2(.,.;\rho)$ be the distribution function of the bivariate normal random vector with mens zero, variances 1 and off-diagonal elements of the 2×2 covariance matrix $\mathbf{R}$ equal to $\rho \in (-1, 1)$. Then the *Gaussian copula* is given by

$$C(v_1, v_2; \rho) = \Phi_2(\Phi^{-1}(v_1), \Phi^{-1}(v_2); \rho), \quad v_1, v_2 \in (0, 1), \tag{3}$$

where $\Phi(.)$ is the distribution function of a standard normal random variable. By Sklar's theorem, for any two marginal distribution functions $F_1(.)$ and $F_2(.)$, the distribution defined as

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2); \rho) = \Phi_2(\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2)); \rho)$$

is a bivariate distribution function whose marginals are $F_1(y_1)$ and $F_2(y_2)$ respectively, and the copula that connects $F(y_1, y_2)$ to $F_1(y_1)$ and $F_2(y_2)$ is the Gaussian copula (3).

Hence Sklar's theorem allows one to construct bivariate distributions with non-Normal marginal distributions and Gaussian copula. Notice that if Y_1 and Y_2 are normally distributed, the resulting joint density generated with the Gaussian copula is reduced to the usual bivariate normal distribution.

Song (2000) shows that the density of the Gaussian copula (3) is

$$\exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{R}^{-1}\mathbf{y} + \frac{1}{2}\mathbf{y}^T\mathbf{y}\right) |\mathbf{R}|^{-\frac{1}{2}}, \tag{4}$$

where $\mathbf{y} = (y_1, y_2)$.

Archimedian copula. Archimedian copulas take a general form

$$C(v_1, v_2) = \phi(\phi^{-1}(v_1) + \phi^{-1}(v_2)),$$

where $\phi : [0, 1] \to [0, \infty]$ is a strictly decreasing function such that $\phi(1) = 0$, called *generator function*. For example, $\phi(t) = (1 + t)^{-1/\theta}$ generates the *Clayton copula*, $\phi(t) = \log(\frac{\theta^t - 1}{\theta - 1})$ generates the *Frank copula*, etc. See Nelsen (2006), Table 4.1, for a summary of Archimedian generator functions.

1.2 Discrete case

It is possible to derive a unique copula presentation for every continuous multivariate distribution, but the same is not true for discrete random variables. Let the distribution of a discrete random variable Y is given by $P(Y = y_r) = p_r$, $r = 1, 2, \ldots$ and $\sum_{r=1}^{\infty} p_r = 1$. Then $P(F_Y(Y) \leq u) \leq u$, $u \in (0, 1]$ and the distribution of $F_Y(Y)$ is given by

$$P(F_Y(Y) = p_1 + \cdots + p_k) = p_k, \quad k = 1, 2, \ldots,$$

which is discrete, and therefore faraway from the uniform distribution on (0,1). The problem is, that $F_Y(Y)$ is one-to-many, while $F_Y^{-1}(Y)$ is many-to-one functions, see Remark 3 also.

Let us note that a version of Sklar's Theorem is offered recently by Niewiadomska-Bugaj and Kowalczyk (2005) who find a unique copula $C_U(.,.)$ in the case when the marginal distributions $F_1(.)$ and $F_2(.)$ can be of any type (continuous, discrete, mixed) and for which the joint distribution function can be defined. The intermediate technique used is the introduced by Szczesny (1991) the so-called *Grade transformation*, which is an extension of the probability integral transformation in continuous case.

The lack of uniqueness of copula presentation for discrete distributions is theoretical issue, but does not inhibit empirical applications. Researchers use copulas because they do not know the joint distribution, so whether working with continuous or discrete data, a pivotal modelling problem is to choose a copula that adequately captures dependence structures of the data without sacrificing attractive properties of marginals.

Applying copulas to discrete data is not as straightforward as in the case of continuous data. Interpreting the dependence parameter of a copula for discrete distribution can be difficult. For continuous copulas, the dependence parameter θ is usually converted to measures such a Kendall's tau (τ_C) or Spearman's rho (ρ_C) defined by

$$\tau_C = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 \quad \text{and} \quad \rho_C = 12\int_0^1\int_0^1 C(u,v)dudv - 3.$$

Both measures are bounded on the interval $[-1, 1]$ and both do not depend on the functional forms of the marginal distributions. For discrete data, however, Marshall (1996) explains that the usefulness of both measures is problematic because they depend (i.e. are not invariant) on the choice of marginal distributions. Tiit and Kaarik (1996), Joe (1997, Section 3) and Van Ophem (1999) are examples of studies that focus explicitly on copula-based models for discrete data.

Example 3 (Association parameter in the Frank copula). An important feature of Frank model is that, it can easily deal with non-overlapping clusters composing a random vector, since the corresponding joint distribution has the same for any cluster size.

For continuous variables, the association parameter θ (which has the properties of within-cluster association parameter) is independent of the marginal distributions and is closely related to the Kendall's τ_C, i.e. $\tau_C = 1 + 4[D_1(\theta) - 1]\theta^{-1}$ where $D_1(\theta) = \int_0^\theta \frac{t}{\theta(\exp(t)-1)}dt$ is the Debye function of order 1, see Magnus et al. (1966). For binary variables, such a dependence is, in general, not true.

However, it can be shown that the relationship between θ and the pair-wise odds-ratio (OR_{ij}) between any two binary variables Y_i and Y_j with marginal probabilities $P(Y_i = 1) = p_i$ and $P(Y_j = 1) = p_j$, $i \neq j$, is as follows:

$$OR_{ij} = \frac{P(Y_i = 1, Y_j = 1)P(Y_i = 0, Y_j = 0)}{P(Y_i = 1, Y_j = 0)P(Y_i = 0, Y_j = 1)} = \frac{(p_i + p_j - 1 - \Delta_{ij})(-\Delta_{ij})}{(1 - p_i + \Delta_{ij})(1 - p_j + \Delta_{ij})},$$

with

$$\Delta_{ij} = \frac{1}{\theta} \, log \left\{ 1 + \frac{[\exp(-\theta(1 - p_i)) - 1][\exp(-\theta(1 - p_j)) - 1]}{\exp(-\theta) - 1} \right\}.$$

Thus the correspondence between the OR_{ij} and θ can be derived for different values of (p_i, p_j).

Remark 2 (Regression estimation using copulas with discrete marginals). As in Remark 1, each marginal discrete distribution can be specified to be conditioned on a vector of covariates. Estimation proceeds by first selecting the appropriate copula $C(.,.;\theta)$ and marginal distributions $F_1(y_1|x_1, \beta_1)$ and $F_2(y_2|x_2, \beta_2)$ where x_1 and x_2 are covariates, and β_1 and β_2 are unknown parameters. In this case, the copula density does not satisfy (2), but it is formed by taking second order differences. Suppressing conditioning on x_1 and x_2 one should use the relation

$$c(F_1(y_{1i}), F_2(y_{2i})) = C(F_1(y_{1i}), F_2(y_{2i}))$$

$$-C(F_1(y_{1i} - 1), F_2(y_{2i})) - C(F_1(y_{1i}), F_2(y_{2i} - 1)) + C(F_1(y_{1i} - 1), F_2(y_{2i} - 1)). \quad (5)$$

The corresponding log-likelihood function is maximized (with respect to β_1, β_2 and θ) using a quasi-Newton iterative algorithm requiring only first derivative. After estimating parameters simultaneously, variances of estimates can be obtained using the robust sandwich formula, for example.

2 The Gaussian copula regression model

The density of p-variate Gaussian copula is given by (4) when $\mathbf{Y} = (Y_1, \ldots, Y_p)$ is a $p-$variate normal vector with zero mean and covariance matrix $\mathbf{R}$, i.e. $\mathbf{Y} \sim N_p(0, \mathbf{R})$

Suppose there are n observations $\mathbf{y}_1, ..., \mathbf{y}_n$, each of dimension p. The *Gaussian copula regression model* can be defined as

$$y_{ij} = h_{ij}^{-1}(z_{ij}), \quad j = 1, \ldots, p, \quad \mathbf{z}_i \sim N_p(0, \mathbf{R}) \quad (6)$$

for $i = 1, \ldots, p$, see Pitt et al. (2006). In (6), $h_{ij}^{-1}(.) = F_{ij}^{-1}(\Phi(.))$, where $F_{ij}^{-1}(.)$ are the univariate distribution function of a continuous or discrete random variable.

Furthermore, we suppose that

$$F_{ij}(.) = F_j(.;\theta_j, x_{ij}), \quad i = 1, \ldots, n,$$

which means that the marginal distribution of the j–th component is the same for all cases, and depends on the vector parameter θ_j and $m \times 1$ vector of covariates x_{ij}. Typically we might have a marginal generalized linear model for a given y_{ij}.

The above approach assumes that the marginal distribution of variables are specified, and uses latent variables to transform each of marginals to a standard Normal distribution. The dependence structure between original variables is created by assuming a multivariate Gaussian distribution for the latent variables. The multivariate Probit Model, e.g. Chib and Greenberg (1998), is a simple example of a Gaussian copula, with univariate probit regressions as the marginals. Oakes and Ritz (2000) consider a bivariate copula regression model with identical marginals whose parameters are known, providing a method for estimating the parameters of discrete marginals.

In the applications, we write $\theta_j = (\beta_j', \psi_j')'$, where β_j' is $m \times 1$ coefficient vector of x_{ij}, and ψ_j is a vector of all other parameters associated with component j.

Example 4 (Multivariate Capital Asset Pricing Model). The Capital Asset Pricing Model is used in finance to quantify the trade-off between the expected risk and return of an investment. It is known that the market returns exhibit systematic deviations away from normality, displaying higher peaks and heavier tails than allowed by the normal distribution. We model the multivariate Capital Asset Pricing Model using the Gaussian copula that has t-distributed marginals, each having its degrees of freedom parameter. The model specification is

$$y_{ij} = x_i\beta_j + \sigma_j e_{ij}, \quad i = 1, \ldots, n;\ j = 1, \ldots, p,$$

where y_{ij} is the excess return of the j-th stock at time i; x_i is the excess market return at time i and the errors (which do not depend on observations) are

$$e_{ij} \sim t_{\nu_j}\sqrt{\frac{\nu_j - 2}{\nu_j}},$$

where t_{ν_j} is a t-distribution with ν_j degrees of freedom, and $\nu_j > 2$ for all j. The errors are standardized to have a variance 1, making it possible to compare the estimated values of the variances σ_j^2.

The dependence of the errors $e_i = (e_{i1}, \ldots, e_{ip})$, being latent variables, are modelled through Gaussian copula. The parameter vector for the j-th equation is $\theta_j = (\beta_j, \sigma_j^2, \nu_j)$.

In fact, we assume that the marginal density of the j-th component is a t-distribution with mean $\mu_{ij} = x_{ij}'\beta_j$, and the parameter ψ_j in formulation (6) is its degrees of freedom parameter ν_j.

Observation 1 (Pitfalls of the correlation coefficient). The key statistic in a bivariate Gaussian distribution is the correlation coefficient. Usually, a low correlation coefficient between two markets implies a good opportunity for an investor to diversify his investment risk. Hence, based on Gaussian assumption, an investor can significantly reduce his risk by balancing his portfolio with investments in the foreign market.

Embrechts at al. (2002) discussed some pitfalls of the usual correlation coefficient as a global measure of dependence, (in the case of high variances, for example). As we noted, the Gaussian copula represents the dependence structure of a joint normal distribution, but a Gaussian copula does not necessarily imply a joint normal distribution unless the marginals are also normal.

There is increasing evidence indicating that Gaussian assumptions are inappropriate in the real world. It has been found that correlations computed with different conditions could differ dramatically, see for example Ang et al. (2002) who studied the correlations between a portfolio and the market conditional on downside movements. It has been found that correlations conditional on large movements are higher than that conditional on small movements. This phenomena has also been characterized as "correlation breakdown", and it is widely discussed in the literature.

Boyer et al. (1999) proposed that in situations of "correlation breakdown", the correlations can reveal little about the underlying nature of dependence, see also the approach suggested by Engle (2002). Therefore, although conditional correlations provide more information about the dependence than the usual Pearson's correlation, the results are sometimes misleading and need to be interpreted very carefully.

Remark 3 (Difference between discrete and continuous models). Although the model (6) is the same for both discrete and continuous components y_{ij},

- the h_{ij} and h_{ij}^{-1} are one-to-one functions for a continuous component y_{ij}

- but h_{ij} is one-to-many functions for a discrete component, with h_{ij}^{-1} a many-to-one function.

This difference between the continuous and discrete models implies that covariates x_{ij} are not observed for a discrete component, but has to be generated in the simulation (if using a Bayesian approach). The priors of the parameter vector Θ are standard, but the prior of $\mathbf{R}$ is more complex, as shown by Pitt et al. (2006).

3 Transition regression models based on copulas

A number of transition regression models for non-Gaussian responses have been proposed in literature, see Benjamin et al. (2003) for a review. Many of these models can be criticized that their semi-parametric formulations (such as a quasi-likelihood and generalized estimating equations) do not lend them self readily to statistical inference and hypothesis testing. A likelihood-based methodology is attractive, but serious problem is the lack of tractable conditional distributions for non-Gaussian responses for either time series or longitudinal data.

3.1 Construction of first order copula transition models

To establish context, we briefly review the first order regressive model for normal responses. Consider the stationary time-series $\{Y_t, t = 1, 2, \ldots\}$ with marginal responses $Y_t \sim N(\beta^T \mathbf{x}_t, \sigma^2)$ for $t = 1, 2, \ldots$. Thus, $\beta^T \mathbf{x}_t$ is the marginal mean of Y_t, $\mathbf{x}_t$ is a vector of explanatory variables observed at time t, β is the corresponding vector of unknown regression coefficients and σ^2 is the variance of the marginal response. If the correlation between lagged responses Y_{t-1} and Y_t is $\rho \in (-1, 1)$, the transition model has the following specification

$$Y_t | Y_{t-1} \sim N(\beta^T \mathbf{x}_t + \rho[Y_{t-1} - \beta^T \mathbf{x}_{t-1}]; \sigma^2(1-\rho^2)). \tag{7}$$

If the state space is continuous, the joint distribution $F(y_1, y_2)$ has two univariate marginal distributions, both equal to the stationary distribution $H(y)$ with a density $h(y)$. The transition distribution $F_{2|1}(y_2|y_1) = P(Y_2 \le y_2 | Y_1 = y_1)$ can be computed as

$$F_{2|1}(y_2|y_1) = \frac{\partial F(y_1, y_2)/\partial y_1}{\partial H(y_1)/\partial y_1}.$$

Although there are many bivariate distributions in literature, few of these share the property of having the same marginal distributions. As we noted, the copula model has the appealing feature of being independent of its marginals. Applying (2), the corresponding conditional density function $f_{2|1}(y_2|y_1)$ can be written using the copula density function as

$$f_{2|1}(y_2|y_1) = h(y_2) c(H(y_1), H(y_2)). \tag{8}$$

If one wishes to obtain a first-order Gaussian autoregressive model in terms of copula transition model corresponding to (7), he simply needs to substitute in (8) the Gaussian copula density function $c(.,.;\rho)$ from (4), the standardized margin distribution function $H(y_t) = \Phi[(y_t - \beta^T \mathbf{x}_t)/\sigma]$ and its density.

Remark 4 (First-order transition models for discrete responses). If $h(y_t) = P(Y_t = y_t)$ represents the marginal distribution of Y_t, the family of transition distributions of $\{Y_t\}$ can be characterized by using the discrete bivariate copula and the discrete distribution function $H(y_t) = \sum_{x \le y_t} h(x)$ as follows:

$$F_{2|1}(y_t|y_{t-1}) = \frac{C(H(y_{t-1}), H(y_t)) - C(H(y_{t-1} - 1), H(y_t))}{h(y_{t-1})}.$$

It follows that the corresponding transition density function, defined as

$$f_{2|1}(y_t|y_{t-1}) = P(Y_t = y_t | Y_{t-1} = y_{t-1}) = \frac{C_{2|1}(y_t, y_{t-1})}{h(y_{t-1})},$$

where the numerator $C_{2|1}(,.,)$ should be calculated as in (5).

In practice, it is possible to extend the transition copula models described above to higher-order representations by mixture transition distribution, e.g. Escarela et al. (2006).

3.2 Semiparametric copula regression models

In economic and financial applications, estimating the dependence parameter is not the ultimate aim. One is more interested in estimating or forecasting certain features of the transition distribution of the time series such as the conditional moment and conditional quantile functions. For example, estimating the conditional value-at-risk of portfolio of assets, or equivalently the conditional quantile of portfolio of assets, has become routine in risk management, see e.g., Engle and Manganelli (2007).

This can be easily accomplished for copula-based semiparametric time series models as the transition distribution of this class is completely characterized by the marginal distribution and copula dependence parameter, see (8). The marginal distribution characterizes the marginal behavior such as the fat-tailedness of the time series $\{Y_t\}$, while the copula function characterizes the temporal dependence property such as non-linear, asymmetric dependence of time series.

Let $\{Y_t\}$ be a stationary Markov process of order one. Here we will assume that the marginal distribution $G^*(.)$ is left unspecified, but the copula function has a parametric form. The function $G^*(.)$ can be non-parametrically estimated by the empirical distribution, its rescaled version, i.e. $\frac{1}{n+1}\sum_{t=1}^{n} I\{Y_t \leq y\}$, or by using standard kernel estimators. If the marginal distribution $G^*(.)$ belongs to a parametric class of distributions, the corresponding stationary Markov processes was studied by Joe (1997). Following (8), we can write the conditional density of Y_t given Y_{t-1} by

$$f^*_{2|1}(Y_t|Y_{t-1}) = g^*(Y_t)c(G^*(Y_t), G^*(Y_{t-1}); \theta),$$

where $g^*(.)$ is the density of the marginal distribution $G^*(.)$ which is unspecified.

One obvious advantage of the copula approach is to separate the temporal dependence structure from the marginal behavior. This is practically important when it is known that the dependence structure and the marginal properties of the time series are affected by different exogenous variables, which can be easily modeled by the copula approach by letting copula parameter θ depending on x_{1t}, say, and the marginal distribution $G^*(.)$ depend on z_{2t}, which may differ from x_{1t}. A related advantage is that the copula measure of temporal dependence is invariant to any increasing transformation of the time series.

Observation 2. The transformed process, $\{U_t : U_t = G^*(Y_t)\}$, is a stationary parametric Markov process. If the joint distribution of U_t and U_{t-1} is given by the copula $C(u_0, u_1; \theta)$, then the conditional density of $U_t|U_{t-1} = u_0$ is just the copula density $c(u, u_0; \theta)$.

As discrete-time Markov models in econometrics are typically expressed as regression models, we will provide in the next two examples such representations for the copula-based stationary Markov time series models, e.g. Chen and Fan (2006).

Example 5 (Semiparametricm model generated by the Gaussian copula). Let the copula $C(.,.;\rho)$ be the Gaussian copula (3). Then the process $\{\Phi^{-1}(G^*(Y_t)\}$ is a Gaussian process that can be represented by

$$\Phi^{-1}(G^*(Y_t)) = \rho\Phi^{-1}(G^*(Y_{t-1})) + \epsilon_t, \tag{9}$$

where $\epsilon_t \sim N(0, 1-\rho^2)$ and ϵ_t is independent of Y_{t-1}. We distinguish the following particular cases of (9):

- if $G^*(Y_t)$ is the standard normal distribution, then $\{Y_t\}$ is a linear AR(1) process;

- if $G^*(Y_t)$ is left unspecified, then we have the class of semiparametric model generated by the Gaussian copula;

- by allowing $G^*(.)$ to be non-normal (Student's t, for example), we obtain the first order Markov process characterized by the Gaussian copula, but with non-normal marginal distributions.

Example 6 (Semiparametric regression transformation models). These are the models defined by

$$\Lambda_{1,\theta_1}(G^*(Y_t)) = \Lambda_{2,\theta_2}(G^*(Y_{t-1})) + \sigma_{\theta_3}(G^*(Y_{t-1}))e_t, \tag{10}$$

where

- $G^*(Y_t)$ is the unknown distribution function of Y_t,

- $\Lambda_{1,\theta_1}(.)$ is a parametric increasing function,

- $\Lambda_{2,\theta_2}(.)$ and $\sigma_{\theta_3}(.) > 0$ are also parametric functions,

- e_t is independent of Y_{t-1} and $\{e_t\}$ is i.i.d. with a parametric density $h(.,\theta_4)$ having mean zero and variance 1.

It is easy to see that in this case $\{Y_t\}$ in (10) is generated by the copula density

$$c(u_0, u_1, \theta) = h\left(\frac{\Lambda_{1,\theta_1}(u_1) - \Lambda_{2,\theta_2}(u_0)}{\sigma_{\theta_3}(u_0); \theta_4}\right) \times \frac{\partial\Lambda_{1,\theta_1}(u_1)}{\partial u_1},$$

where θ consists of the distinct elements $\theta_1, \theta_2, \theta_3$ and θ_4.

The stationary Markov process with the Gaussian copula in Example 5 and a non-parametric marginal distribution $G^*(.)$ can be obtained by substituting in (10): $\Lambda_{1,\theta_1}(u_1) = \Phi^{-1}(u_1)$, $\Lambda_{2,\theta_2}(u_0) = \rho\Phi^{-1}(u_0)$, $\sigma_{\theta_3}(u_0) = \sqrt{1-\rho^2}$, $h(.,\theta_4)$ is the standard normal density, and $\theta = \theta_1 = \theta_2$.

A generalizations of the model in Example 6 (free of independence restriction between error term e_t and Y_{t-1}) is considered by Chen and Fan (2006) where is shown how copula-based time-series specifications lead to semiparametric quantile regression models. The authors proposed simple estimators of the unknown marginal distribution and copula dependence parameters, which serve as a base for estimators of conditional moments and conditional quantiles and their large sample properties are discussed for Gaussian, Frnk and Clayton copulas.

4 A copula based regression model for unobserved heterogeneity

It is well known that the majority of private health insurance coverage in the English-speaking world is financed by the employers. Families with two working spouses might be confronted with more insurance choices than if only one spouse is employed. The implication is that spouses must decide whether to enroll together in the same insurance plan, or they might acquire separate plans.

The question of empirical interest is whether enrollment in separate plans and spousal health care use are jointly determined, i.e. if a husband's and wife's utilization and their insurance choice are simultaneously determined.

Table 1. Sample Means by Insurance Arrangement

		Enrolled in same plan, N=5250		Enrolled in different plans, N=1386	
		Wives	Husbands	Wives	Husbands
Utilization					
PV	No. Physician visits	3.92	2.33	4.25	2.36
NPV	No. Nonphysician visits	1.70	1.03	1.87	0.99
ER	No. of Emergency Room visits	0.11	0.10	0.10	0.10
Employment					
EMPLOYED	= 1 if employed	0.77	0.96	0.96	0.97
SELFFEMP	= 1 if self employed	0.11	0.17	0.05	0.09

Table 1 presents variable definitions and summary statistics (data come from 1996-1999 data-base of the Medical Expenditure Panel Study in USA). For example, it can be seen that

– the wives tend to be more intensive users of health care than husbands;

– spouses in separate plans appear to have slightly more physician visits and a similar number of nonphysician visits and emergency visits compared to couples enrolled in the same plan;

– spouses enrolled in separate plans have a higher proportion of employed wives and lower rates of self-employment for both spouses, etc.

4.1 Regression specifications of the model

The wife and husband in family i have a latent disposition to use the health care services denoted as $Y_{i,w}^U$ and $Y_{i,h}^U$ and the couple's latent tendency to enroll in separate plans is denoted by D_i^U. The equations underlying these latent variables are assumed to be linear:

$$Y_{i,w}^U = \mathbf{x^T}_{i,w}\beta_w + u_{i,w} + \lambda_{v,w}v_i, \tag{11}$$

$$Y_{i,h}^U = \mathbf{x^T}_{i,h}\beta_h + u_{i,h} + \lambda_{v,h}v_i \tag{12}$$

and

$$D_i^U = \mathbf{z^T}_i\alpha + \lambda_u(u_{i,w} + u_{i,h}) + \epsilon_i. \tag{13}$$

The quantities in (11), (12) and (13) have the following meaning:

- $\mathbf{x^T}_{i,w}$ and $\mathbf{x^T}_{i,h}$ are vectors of explanatory variables that affect the wife's and husband's utilization, respectively;
- the vector $\mathbf{z^T}_i$ consists of variables that affect the insurance choice;
- β_w, β_h and α are coefficient to be estimated;
- λ-coefficients reflect the underlying data dependence structure: λ_u in (13) measure how the insurance choice is related to family health consumption, whereas $\lambda_{v,w}$ and $\lambda_{v,w}$ in (11) and (12) indicate the degree of wives and husbands contribution to utilization;
- ϵ_i are independently distributed error terms, assumed to be uncorrelated with $(u_{i,w}, u_{i,h}, v_i)$;
- u_i and v_i represent the dependence between spouses' utilization and insurance choice and are not observed by the researcher;

The terms $u_{i,w}$ and $u_{i,h}$ represent unobserved characteristics that affect insurance choice as well as each spouse's utilization. In a similar manner, the term v_i represents the unobserved family trait, independent from those contained in u_i that causes the spouses' utilization to be correlated (for example, wives and husbands might share similar diets and lifestyles that affect spousal health care consumption).

In the equations (11), (12) and (13) the outcome variables $Y_{i,w}^U$, $Y_{i,h}^U$ and D_i^U are unobservable latent tendencies with corresponding observable variables that can be studied. The observable utilization variables $y_{i,w}$ and $y_{i,h}$ are recorded as discrete counts of the visits to certain health care providers. The wife's utilization probability for a particular medical service is assumed to follow a negative binomial (NB) distribution in the form

$$f_1(y_{i,w}|\mu_{i,w}) = \frac{\Gamma(y_{i,w}+\delta)}{\Gamma(\delta)\Gamma(y_{i,w}+1)}\left(\frac{\delta}{\lambda_{i,w}+\delta}\right)^{\delta}\left(\frac{\lambda_{i,w}}{\lambda_{i,w}+\delta}\right)^{y_{i,w}}, \tag{14}$$

where $\mu_{i,w} = exp(\mathbf{x^T}_{i,w}\beta_w)$ is the conditional mean, and $\delta^{-1} > 0$ is the overdispersion parameter in the conditional variance $\mu_{i,w}(1 + \delta\mu_{i,w})$. This functional form provides substantial modelling flexibility. For the husband we use the similar to (14) NB probability function $f_2(y_{i,h}|\mu_{i,h})$.

The variable D_i is dichotomous variable indicating whether the husband and wife in family i are enrolled in separate insurance plans. Couple i's decision is assumed to follow a Probit specification for which the contribution to the unlogged likelihood function L_D is

$$L_D = [\Phi(\mathbf{z}^{\mathbf{T}}{}_i\alpha)]^{D_i}[1 - \Phi(\mathbf{z}^{\mathbf{T}}{}_i\alpha)]^{1-D_i},$$

where $\Phi(.)$ is the standard normal distribution function.

The empirical model considers three sources of utilization: physician visits (PV), nonphysician visits (NPV) and emergency room visits (ER). For each of them we estimate the unknown parameters under the following assumptions: The composite error terms corresponding to outcome variables in (11), (12) and (13) are

$$\eta_{i,w} = u_{i,w} + \lambda_{v,w}v_i, \ \eta_{i,h} = u_{i,h} + \lambda_{v,h}v_i \quad \text{and} \quad \eta_{D,i} = \lambda_u(u_{i,w} + u_{i,h}) + \epsilon_i,$$

correspondingly. If u_w, u_h, v and ϵ_i are independent and $N(0,1)$ distributed, then the covariances of error terms are

$$Cov(\eta_w, \eta_h) = \lambda_{v,w}\lambda_{v,h}, \ Cov(\eta_w, \eta_D) = \lambda_u \quad \text{and} \quad Cov(\eta_h, \eta_D) = \lambda_u.$$

However, the latent variables u_w, u_h and v are unobserved. Because the variances of the latent factors cannot be identified, a normalization on either $\lambda_{v,w}$ or $\lambda_{v,h}$ is required. In results discussed below, $\lambda_{v,w}$ is normalized to unity.

In order to estimate the λ-coefficients (corresponding to θ-parameters describing the dependence structure given by copula model), random values of the three latent variables (u_w, u_h and v) are drawn from independent standard normal distributions. The simulated values (u_w^S, u_h^S, v^S) are treated as observed variables in the joint density function of the three outcomes. The process is repeated S times, and the average of the S draws provides a simulated density

$$f^*(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}) = \frac{1}{S}\sum_{s=1}^{S} f(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, u_{w,s}^S, u_{h,s}^S, v_s^S).$$

For maximization of the likelihood function based on that simulated density we use the maximum simulated likelihood (MSL) technique, see Train (2003) for details. The results are shown in the bottom panel of Table 2.

4.2 Trivariate Frank copula specification

The copula's facility to accommodate unobserved heterogeneity is especially important for the data analyzed. Joint estimation to single equation estimation is preferable because variation in utilization and insurance choice, due to the heterogeneity in $(u_{i,w}, u_{i,h}$ and $v_i)$, should be captured by the dependence parameters of appropriate trivariate copula. The two first marginals $F_1(.)$ and $F_2(.)$ are wife's NB and husband's NB distributions (14), respectively. The third marginal of the corresponding copula, $F_3(.)$, is the Probit model. Since all marginals are discrete, the corresponding joint copula probability mass function is obtained by the difference equation

$$c(F_1(y_{1i}), F_2(y_{2i}), F_3(y_{3i})) = C(F_1(y_{1i}), F_2(y_{2i}), F_3(y_{3i}))$$

$$-C(F_1(y_{1i}-1), F_2(y_{2i}), F_3(y_{3i})) - C(F_1(y_{1i}), F_2(y_{2i}-1), F_3(y_{3i}))$$

$$-C(F_1(y_{1i}), F_2(y_{2i}), F_3(y_{3i}-1)) + C(F_1(y_{1i}-1), F_2(y_{2i}-1), F_3(y_{3i}))$$

$$+C(F_1(y_{1i}), F_2(y_{2i}-1), F_3(y_{3i}-1)) + C(F_1(y_{1i}-1), F_2(y_{2i}), F_3(y_{3i}-1))$$

$$-C(F_1(y_{1i}-1), F_2(y_{2i}-1), F_3(y_{3i}-1)),$$

which extends (5) in three dimensions.

Remark 4 (Trivariate mixture of powers copula representation). Let $M(.)$ be a "mixing" distribution function with Laplace transform $\phi(s), s > 0$. For any distribution function $A(x)$ and mixing function $M(.)$, there exists a unique distribution function $B(x)$ satisfying

$$A(x) = \int_0^\infty B(x|\theta_1)dM(\theta_1) = \phi(-\log B(x)).$$

The term $\theta_1 > 0$ can be thought of as unobserved heterogeneity that affects x. The choice of $M(.)$ determines the functional form of the Laplace transformation $\phi(.)$, which, in turn, determines the functional form of $A(x)$. Solving the last equation for $B(.)$ gives $B(x) = exp(-\phi^{-1}(A(x))$, e.g. Joe (1993).

Since bivariate copula is a distribution function, it can be expressed in terms of mixtures powers as

$$C(u, v, \theta_1) = \int_0^\infty B(u|\theta_1)B(v|\theta_1)dM(\theta_1) = \phi(-\log B(u) - \log B(v)),$$

i.e. $C(u, v, \theta_1) = \phi(\phi^{-1}(A(u)) + \phi^{-1}(A(v)))$ and the heterogeneity term θ_1 affects both u and v, see Joe (1997). If $A(.)$ is uniformly distributed on $(0, 1)$, i.e. $A(u) = u$ for $u \in (0, 1)$, then the last expression takes the bivariate Archimedian form given at the end of Section 1.1.

The trivariate mixtures of powers representation is

$$C(u, v, w, \theta_1, \theta_2) = \int_0^\infty \int_0^\infty B(u|\theta_2)B(v|\theta_2)dM_2(\theta_2; \theta_1)B(w|\theta_1)dM_1(\theta_1) \tag{15}$$

In this formulation, the heterogeneity term θ_1 affects u, v and w and the second heterogeneity term θ_2 affects u and v. Representation (15) is symmetric with respect to (u, v) but not with respect to w. More precisely, the parameter θ_2 measures dependence between u and v. The parameter θ_2 measures the dependence between u and w as well as between v and w, and the two must be equal. Thus, the copula representation (15) accommodates unobserved heterogeneity in the tree outcomes.

The distribution of $M_1(.)$ has Laplace transform $\psi(.)$ and $M_2(.)$ has Laplace transform $[(\psi^{-1}\circ\phi)_{-1}(-\theta_1\log(.))]^{-1}$, where $\psi\circ\phi$ denotes the functional operation $\psi(\phi(.))$.

When $\psi = \phi$, (15) simplifies to

$$C(u, v, w, \theta_1) = \psi(\psi^{-1}(u) + \psi^{-1}(v) + \psi^{-1}(w)),$$

i.e. the trivariate Archimedian copula which is symmetric in the sense that produces one dependent parameter θ_1 affecting u, v and w.

When $\psi \neq \phi$, (15) can be written as

$$C(u, v, w, \theta_1, \theta_2) = \psi(\psi^{-1}\circ\phi[\phi^{-1}(u) + \phi^{-1}(v)] + \psi^{-1}(w)).$$

Therefore, different functional forms of the Laplace transformation produce different families of trivariate copulas. For higher dimensions see McNeil (2007).

Substituting $\phi(s) = \exp(-s^{1/\theta_1})$ and $\psi^{-1}\circ\phi(s) = s^{\theta_1/\theta_2}$ in (15) we obtain the *trivariate Frank copula*

$$C_F(u, v, w, \theta_1, \theta_2) = -\theta_1 log\left\{1 - \left(1 - [1 - c_2^{-1}(1 - e^{-\theta_2 u})(1 - e^{-\theta_2 v})]^{\theta_1/\theta_2}\right) \frac{1 - e^{-\theta_1 w}}{c_1}\right\}, \tag{16}$$

where $c_1 = 1 - e^{-\theta_1}$, $c_2 = 1 - e^{-\theta_2}$ and $0 < \theta_1 \leq \theta_2$, e.g. Joe (1993). The copula given by (16) is an appropriate model for the unobserved heterogeneity of insurance data considered. The thrivariate Frank copula appears to be more stable in dealing with large count values than Clayton copula, for example (which is also a particular case of (15)), e.g. Zimmer and Trivedi (2006).

The trivariate Frank copula (16) respects the dependence structure between the three marginals for our data (the wife's health care demand; the husband's health care demand and the family insurance agreement choice) by permitting two types of interdependence.

The copula approach has two advantages. First, the formulation of copula approach allows for a measure of family selection into insurance status (for example, if a family anticipates particular health care needs in the future, it might choose an insurance arrangement that provides generous coverage of appropriate services). Second, the copula approach allows for a measure of interdependence between spouses' health care demands based on unobserved family heterogeneity.

The dependence parameter θ_1 measures the degree to which the family's insurance agreement decision is related to its health care utilization. This relation is decomposed of two separate effects. First, as discussed above, θ_1 includes the indirect selection effect of being enrolled in different plans. In addition, θ_1 also measures the direct casual effect on utilization of being enrolled in different plans.

The joint density $f(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, u_w, u_h, v)$ for the trivariate Frank copula (16) can be decomposed by Bayes' rule

$$f(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, u_w, u_h, v) = f_{12|3}(y_w, y_h|\mathbf{x}_w, \mathbf{x}_h, D, v) \times f_3(D|\mathbf{z}),$$

where u_w and u_h no longer appear in the right hand because utilization is conditional on D. Therefore, the distribution of utilization conditional on insurance choice is expressed as follows

$$f_{12|3}(y_w, y_h|\mathbf{x}_w, \mathbf{x}_h, D, v) = \frac{f(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, u_w, u_h, v)}{f_3(D|\mathbf{z})} \tag{17}$$

For the data considered and for each of utilization measures (physician visits (PV), nonphysician visits (NPV) and emergency room visits (ER), as in Section 4.1), the numerator in (17) is estimated by Frank copula (16), and the denominator is estimated by the usual maximum likelihood method applied to the Probit model. Estimates of copula parameters for each of utilization measures are presented in the upper panel of Table 2.

Table 2. Comparison with the MSL Model

	PV		NPV		ER	
	Coeff.	St. Err.	Coeff.	St. Err.	Coeff.	St. Err.
Trivariate Copula Model						
Selection Coefficient (θ_1)	0.025	0.014	0.039	0.083	0.098	0.026
Insurance Choice Coefficient (θ_2)	1.127	0.037	2.273	0.199	1.844	0.252
Log Likelihood	-30713.92		-17364.86		-7396.02	
Simulation-Based Model (MSL)						
Selection Coefficient (λ_u)	0.023	0.024	-0.052	0.042	0.035	0.056
Correlation Coefficient ($\lambda_{v,h}$)	0.438	0.034	0.585	0.082	0.766	0.154
Log Likelihood	-30743.33		-16969.16		-6918.72	

Comparing estimates in the upper and lower panels in Table 2, one can conclude that the quantitative conclusions reveals several similarities by using the regression analysis and copula-based approach. Both models indicate significant dependence between spouses' utilization (measured by θ_2 in the copula model and $\lambda_{v,h}$ in the MSL model). Both show that correlation is largest in magnitude for nonphysician utilization (NPV) followed by ER and physician usage (PV).

Regarding selection (measured by θ_1 in the copula model and λ_u in the MSL model), the MSL model does not reveal significant selection with respect to any of the measures of utilization. Both estimation approaches show that the selection effect is quantitatively small, especially in comparison to correlation between spouses' utilization.

By comparison, the trivariate copula model is the basis for the conventional likelihood function (albeit a complicated form), and convergence is usually achieved in less than 60 minutes. The MSL model used in Section 4.1, requires three times more memory than the copula model, a significant disadvantage when working with large samples.

Despite the fact that D is not included as an explanatory variable in the utilization equations, its direct effect on utilization can be extracted from estimates of θ_1. By substituting estimated coefficients and dividing the two quantities in (17), a representation of the conditional distribution is obtained from which the casual effect of D on utilization is determined. For example, the effect of D on y_w can be calculated by taking expectations with respect to y_w and summing over all possible values of y_w:

$$E(y_w|D, \mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, \Theta^E) = \sum_{y_h=0}^{\infty} \sum_{y_w=0}^{\infty} y_w \frac{f_{12|3}(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, \Theta^E)}{f_3(D|\mathbf{z}, \Theta^E)},$$

where Θ^E represents estimation coefficients.

The wives's average treatment effect (ATE_w) of being enrolled in different plans is estimated as is estimated as

$$ATE_w = E(y_w|D = 1, \mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, \Theta^E) - E(y_w|D = 0, \mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, \Theta^E),$$

where the covariates are set to their sample means. The husband's treatment effect is obtained analogously.

To calculate the standard deviation of ATE_w, a Monte Carlo procedure can be employed in which Θ is drawn from a normal distribution with mean Θ^E and variance equal to the estimated covariance matrix. The results, reproduced by Zimmer and Trivedi (2006), (who replicated the Monte Carlo procedure 500 times) are shown in Table 3.

Table 3.

A. Treatment Effects from the Copula Model

	Wife		Husband	
Utilization Measure	ATE	St. Err.	ATE	St. Err.
Physician Visits	0.040	0.214	0.022	0.110
Nonphysician Visits	0.018	0.072	0.011	0.039
ER Visits	0.004	0.008	0.004	0.008

B. Treatment Effects from the MSL Model

	Wife		Husband	
Utilization Measure	ATE	St. Err.	ATE	St. Err.
Physician Visits	0.031	0.028	0.017	0.017
Nonphysician Visits	-0.001	0.012	-0.0002	0.009
ER Visits	-0.0001	0.001	-0.0002	0.003

It is not surprising that treatment effects are small and relatively widely dispersed. The only link between D and utilization is trough the dependence parameter θ_1, which estimates are small compared to estimates of θ_2. The interpretation is that the extent in which a family's insurance agreement is related to its health care utilization is small relative to the extent to which spouses' utilization are related to each other.

5 Hyperbolic copula-based longitudinal models

Since copulas are concerned primarily with relationships, one can utilize any multivariate distribution to generate a copula. In this section we will propose the use of parametric copula corresponding to the multivariate distribution with generalized hyperbolic marginals in a longitudinal data framework. First, we describe the model and in Section 5.2 we give basic facts about two families of multivariate distributions that may be used used to generate a copula.

5.1 Data and basic distribution assumptions

The data under interest are composed by risk class i (town) and over time t (year). For each observation $\{it\}$ the responses consist of total claims amount S_{it} and claims number N_{it}. We also have random errors e_{it} and town characteristics, described by the vector $\mathbf{x}_{it}$ of explanatory variables. Hence, the data available consist of

$$\{S_{it}, N_{it}, e_{it}, \mathbf{x}_{it},\ t = 1, \dots, T_i,\ i = 1, \dots, n\}.$$

The total claims amount $S_{it} = \sum_{k=1}^{N_{it}} C_{it,k}$, where $C_{it,k}$, $k = 1, 2, \dots$ are claims resulting from individual losses from the same distribution. The random variables N_{it} and S_{it}, $t = 1, \dots, T_i$, are assumed independent among risk classes.

Let (S_{it}, N_{it}) be vector of responses from the i-th town and the t-th time point. We decompose the joint density $f(s_{it}, n_{it})$ of (S_{it}, N_{it}), (see (2)), as

$$f(s_{it}, n_{it}) = f_{1|2}(s_{it}|N_{it} = n_{it})h(n_{it}).$$

For the claims number component density $h(n_{it})$, there are several candidate distributions that readily accommodate the effect of the errors e_{it}. The authors usually use compound Poisson distributions for claims number (including as special cases the Poisson, negative binomial and Poisson-Inverse-Gaussian distributions). For the severity component, $f_{1|2}(s_{it}|N_{it} = n_{it})$, the exponential family of distributions can be applied.

Consider the history of the i-th risk class and let $\mathbf{H}^N_{i,t-1} = \{N_{i,1}, \dots, N_{i,t-1}\}$ and $\mathbf{H}_{i,t-1} = \{S_{i,1}, N_{i,1}, \dots, S_{i,t-1}, N_{i,t-1}\}$ denote the claims number history and claims history up to time $t-1$, $t = 1, \dots, N$. To model the development of claims over time Frees and Wang (2006) assume that:

- Claim severity depends on current claims number, as well as the entire prior history of the claims process, i.e. the distribution of S_{it} is a function of N_{it} and $\mathbf{H}_{i,t-1}$;

- Claims number depends on the prior history of the claims number process but not the claims severity process. That is, the distribution of N_{it} is a function of $\mathbf{H}^{N}_{i,t-1}$.

Under these assumptions, the joint density $f(.)$ of $\{S_{i,1}, N_{i,1}, \ldots, S_{i,T}, N_{i,T}\}$ may be written as

$$f(S_{i,1}, N_{i,1}, \ldots, S_{i,T}, N_{i,T}) = f_{2|1}(S_1|N_1)h(N_1)\prod_{t=2}^{T}[f(S_t|N_t, \mathbf{H}_{i,t-1}) \times f(N_t|\mathbf{H}^{N}_{i,t-1})]$$

$$= \left[h(N_1)\prod_{t=2}^{T} f(N_t|\mathbf{H}^{N}_{i,t-1})\right]\left[f_{2|1}(S_1|N_1)\prod_{t=2}^{T} f(S_t|N_t, \mathbf{H}_{i,t-1})\right]$$

$$= \{\text{frequency distribution}\} \times \{\text{conditional severity distribution}\}$$

For a model of aggregate losses described, the interest is in predicting both the claims number process as well as the claims amount process. In a longitudinal data framework, one encounters data from a cross-section of risk classes with a history of insurance claims available for each risk class. Further, explanatory variables for each risk class over time are available to help explain and predict both the claims number process and claims amount process.

For the claims severity process one can use:

- a generalized linear model for the marginal distributions (to describe the cross-sectional characteristics), conditional of frequency;

- an appropriate parametric copula to model the joint distribution of claims over time.

Frees and Wang (2005) use a Poisson regression model that is conditioned on a sequence of latent variables. These latent variables drive the serial dependencies among claims numbers and their joint distribution over time is represented via $t-$ and Gaussian copulas. Frees and Wang (2005) focus on elliptical class of copulas applied for credibility ratemaking[1]. By the proposed methodology the authors develop a unified treatment of both the continuous claims amount and discrete claims number process. The procedures developed are employed for automobile liability claims for a sample of $n = 29$ towns of Massachusetts considered annual data from 5 years, 1994-1998.

[1]Credibility ratemaking is a technique for predicting future expected claims of a risk class, given past claims of that and related risk classes, thus employing longitudinal data set-up.

5.2 Multivariate distribution models with generalized hyperbolic marginals and their copulas

A subclass of the multivariate generalized hyperbolic (MGH) distributions, namely the hyperbolic distributions has been introduced via so-called variance-mean mixtures of Inverse Gaussian distributions. This subclass does not have hyperbolic distributed marginals, i.e. the subclass is not closed with respect to passing to marginal distributions. Therefore and because of other theoretical reasons, Barndorff-Nielsen (1977) extended this class to the family of MGH distributions. Many different parametric representations of the MGH density functions are provided in literature. We will use the following one.

Definition 1 (MGH distribution). An n-dimensional random vector $\mathbf{Y}$ is said to have a *multivariate generalized hyperbolic (MGH) distribution with location vector* $\mu \in \Re^n$ *and scaling matrix* $\Sigma \in \Re^{n\times n}$, if it has a stochastic representation $\mathbf{Y} \stackrel{d}{=} \mathbf{A}'\mathbf{X}+\mu$ for some lower triangular matrix $\mathbf{A}' \in \Re^{n\times n}$ such that $\mathbf{A}'\mathbf{A} = \Sigma$ is positive definite and $\mathbf{X}$ has a density function of the form ($\mathbf{x} \in \Re^n$):

$$f_{\mathbf{X}}(\mathbf{x}) = C_n \frac{K_{\lambda-n/2}(\alpha\sqrt{1+\mathbf{x}'\mathbf{x}})}{(1+\mathbf{x}'\mathbf{x})^{n/4-\lambda/2}} \exp(\alpha\beta'\mathbf{x}) \quad \text{with} \quad C_n = \frac{\alpha^{n/2}(1-\beta\beta')^{\lambda/2}}{(2\pi)^{n/2}K_\lambda(\alpha\sqrt{1-\beta\beta'})}. \tag{18}$$

$K_\nu(.)$ denotes the modified Bessel-function of the third kind (or MacDonald function) with index ν (e.g., Magnus at al. (1966), p. 65) and the parameter domain is $\alpha > 0, \lambda \in \Re$ and $\parallel \beta \parallel_2 < 1$ ($\parallel \cdot \parallel_2$ denotes the Euclidean norm). The family of n-dimensional multivariate generalized hyperbolic distributions is denoted by $MGH_n(\mu, \Sigma, w)$ with $w = (\lambda, \alpha, \beta)$.

An important property of the above parameterization of the MGH density function (18) is its invariance under affine-linear transformations. For $\lambda = \frac{n+1}{2}$ we obtain *multivariate hyperbolic* density and for $\lambda = -1/2$ the *multivariate inverse Gaussian* density. Hence $\lambda = 1$ leads to hyperbolically distributed one-dimensional marginals.

An MGH distribution belongs to the class of *elliptically contoured distributions* if and only if $\beta = (0, \ldots, 0)'$. In this case the density function of $\mathbf{Y}$ can be represented as

$$f_{\mathbf{Y}}(\mathbf{y}) = |\Sigma|^{-1/2} g[(\mathbf{y}-\mu)'|\Sigma|^{-1}(\mathbf{y}-\mu)], \quad \mathbf{y} \in \Re^n, \tag{19}$$

for some density *generator function* $g(.) : (0, \infty) \to (0, \infty)$. Let us denote the family of n-dimensional elliptically contoured distributions by $E_n(\mu, \Sigma, g)$[2]. For a detailed

[2] The stochastic representation is $\mathbf{Y} \stackrel{d}{=} \mathbf{A}'\mathbf{X} + \mu$, where $\mathbf{X}$ is a m-dimensional spherically distributed random vector, $\mathbf{A} \in \Re^{m\times n}$ with $\mathbf{A}'\mathbf{A} = \Sigma$, and $rank(\Sigma) = m$. According to the stochastic representation of spherically distributions we can write also $\mathbf{Y} \stackrel{d}{=} R_m \mathbf{A}'\mathbf{U}^{(m)} + \mu$, where $\mathbf{A}'\mathbf{A} = \Sigma$ and the random variable $R_m \geq 0$ is independent of the m-dimensional random vector $\mathbf{U}^{(m)}$ which is uniformly distributed on the unit sphere in $\Re^m$.

treatment of elliptically countered distributions see Schmidt (2002). Particular cases of (19) are listed in Appendix 1.

A disadvantage of the $MGH_n(\mu, \Sigma, w)$ family of distributions is that the marginals Y_i of $\mathbf{Y}$ are not mutually independent for some choice of the scaling matrix Σ. In other words, they do not allow the modelling phenomena where the random variables result as the sum of independent random variables. This pitfall is serious since the independence may be a genuine property of the problem for which the stochastic model is sought. Furthermore, in the case of *asymmetry*, (i.e. when $\beta \neq (0, \ldots, 0)$, the covariance matrix is in a complex relation with the matrix Σ.

It can be shown that if $\mathbf{Y} \in MGH_n(\mu, \Sigma, w)$, the mean vector and covariance matrix are given by

$$E(\mathbf{Y}) = \mu + \alpha R_{\lambda,1}(\sqrt{\alpha^2(1 - \beta\beta'})\mathbf{A}'\beta$$

and

$$Cov(\mathbf{Y}) = \alpha R_{\lambda,1}(\sqrt{\alpha^2(1 - \beta\beta'})\Sigma + [R_{\lambda,2}(\sqrt{\alpha^2(1 - \beta\beta'}) - R^2_{\lambda,1}(\sqrt{\alpha^2(1 - \beta\beta'})]\frac{\mathbf{A}'\beta\beta'\mathbf{A}}{1 - \beta\beta'}$$

where $R_{\lambda,i} = \frac{K_{\lambda+i}(\mathbf{y})}{\mathbf{y}^i K_\lambda(\mathbf{y})}$ for $i = 1, 2$, see Schmidt et al. (2006). For the symmetric case $\beta = (0, \ldots, 0)$, $\lambda = 1$, and the above expressions simplify to $E(\mathbf{Y}) = 0$ and $Cov(\mathbf{Y}) = \frac{K_2(\alpha)}{\alpha K_1(\alpha)\Sigma}$.

Schmidt et al. (2006) propose an alternative concept. Instead of $MGH_n(\mu, \Sigma, w)$ family, a distribution is considered which is composed of n independent marginals with univariate generalized hyperbolic distributions with zero location and unit scaling. Such a canonical random vector is then subject to affine-linear transformation. As a consequence, the transformation matrix can be modelled proportionally to the square root of the covariance-matrix inverse even in the asymmetric case. This property holds, for multivariate normal distributions, for instance.

Definition 2 (Multivariate affine generalized hyperbolic (MAGH) distribution). An n-dimensional random vector $\mathbf{Y}$ is said to be *multivariate affine generalized hyperbolic (MAGH) distributed with location vector* $\mu \in \Re^n$ *and scaling matrix* $\Sigma \in \Re^{n \times n}$, if it has a stochastic representation $\mathbf{Y} \overset{d}{=} \mathbf{A}'\mathbf{X} + \mu$ for some lower triangular matrix $\mathbf{A}' \in \Re^{n \times n}$ such that $\mathbf{A}'\mathbf{A} = \Sigma$ is positive definite and the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ consists of mutually independent random variables $X_i \in MGH_1(0, 1, w_i), i = 1, \ldots, n$. In particular the one-dimensional marginals of $\mathbf{X}$ are generalized hyperbolically distributed. The family of n-dimensional multivariate affine generalized hyperbolic distributions is denoted by $MAGH_n(\mu, \Sigma, w)$ with $w = (w_1, \ldots, w_n)$ and $w_i = (\lambda_i, \alpha_i, \beta_i)'$, $i = 1, \ldots, n$.

Observe that $MAGH_n(\mu, \Sigma, w)$ distribution has independent marginals if the scaling matrix Σ equals to the identity matrix $\mathbf{I}$. However, no $MAGH_n(\mu, \Sigma, w)$ distribution belongs to the class of elliptically contoured distributions for dimension $n \geq 2$.

Schmidt et al. (2006) calculate the mean vector and covariance matrix of $\mathbf{Y} \in MAGH_n(\mu, \Sigma, w)$. The corresponding expressions are

$$E(\mathbf{Y}) = \mathbf{A}'e_X + \mu \quad \text{and} \quad Cov(\mathbf{Y}) = \mathbf{A}'\mathbf{C}\mathbf{A}, +\alpha R_{\lambda,1}(\sqrt{\alpha^2(1-\beta\beta')}A'\beta$$

where

$$e_X = (E(X_1), \ldots, E(X_n)) \quad \text{with} \quad E(X_i) = R_{\lambda_i,1}[\sqrt{\alpha_i^2(1-\beta_i^2)}]\alpha_i\beta_i$$

and $\mathbf{C} = (c_{11}, \ldots, c_{nn})$ with

$$c_{ii} = R_{\lambda_i,1}[\sqrt{\alpha_i^2(1-\beta_i^2)}] + \{R_{\lambda_i,2}[\sqrt{\alpha_i^2(1-\beta_i^2)}] - R^2_{\lambda_i,1}[\sqrt{\alpha_i^2(1-\beta_i^2)}]\}\frac{\beta_i^2}{1-\beta_i^2}.$$

It can be seen that the covariance matrix $Cov(\mathbf{Y})$ is proportional to Σ if $\alpha = \alpha_i$, $\beta = \beta_i$ and $\lambda = \lambda_i$ for all $i = 1, \ldots, n$.

Observation 3 (General affine transformations). The consideration of the lower triangular matrix $\mathbf{A}$ in the stochastic representations of Definitions 1 and 2 is essential since any other decomposition of the scaling matrix Σ would lead to a different class of distributions.

Only the elliptically contoured subclass of the MGH distributions is invariant with respect to different decompositions $\mathbf{A}\mathbf{A}' = \Sigma$. In particular, all decompositions of the scaling matrix Σ lead to same distribution since they enter in (18) via the form $\mathbf{A}\mathbf{A}' = \Sigma$.

However in the asymmetric or general affine case this equivalence does not hold anymore. In such a case, the matrix $\mathbf{A}$ can be sought via s singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}'$, where $\mathbf{W}$ is a diagonal matrix having the square roots of eigenvalues of $\Sigma = \mathbf{A}\mathbf{A}'$ on its diagonal and where the matrix $\mathbf{V}$ consists of the corresponding eigenvectors of Σ. The matrices $\mathbf{W}$ and $\mathbf{V}$ are directly determined from Σ whereas the matrix $\mathbf{U}$ might be some arbitrary matrix with orthonormal columns (rotation and flip). However, the most common case, of course, is $\mathbf{U} = \mathbf{I}$. Here the matrix $\mathbf{A}$ is directly computed from Σ utilizing its eigenvectors and eigenvalues. Consequently, every margin of $\mathbf{X}$ determined by the principal components (i.e. eigenvectors) of the covariance matrix Σ: $\mathbf{X} = \mathbf{A}'^{-1}\mathbf{Y} = \mathbf{W}^{-1}\mathbf{V}'\mathbf{Y}$.

The MGH and MAGH copulae (Schmidt at al. (2006). According to Definitions 1 and 2, the MGH and MAGH distributions are represented by affine-linear transformations of random vectors following the a standardized MGH distribution and MAGH distribution, respectively. Leaving the affine-linear transformation aside, we are interested in the dependence structure (copula) of the underlying random vector. In particular we set the scaling matrix $\Sigma = \mathbf{I}$ and $\mu = 0$. Note that the copula of MGH and MAGH distribution does not depend on the location vector, i.e. μ is not a copula parameter.

Let $\mathbf{X} \in MGH_n(0, \mathbf{I}, w)$. Then the copula density function of $\mathbf{X}$ is given by

$$c(u_1, \ldots, u_n) = C_n \frac{K_{\lambda-n/2}(\alpha\sqrt{1+\mathbf{x}'\mathbf{x}})}{(1+\mathbf{x}'\mathbf{x})^{n/4-\lambda/2}} \times \prod_{i=1}^{n} \frac{(1+x_i^2)^{1/4-\lambda/2}}{K_{\lambda-1/2}(\alpha\sqrt{1+x_i^2})} \frac{\exp(\alpha\beta'\mathbf{x})}{\exp(\prod_{i=1}^{n} \alpha\beta_i x_i)}\Big|_{x_i=F_i^{-1}(u_i)}$$

for $u_i \in [0,1]$, $i = 1, \ldots, n$ and some normalizing constant C_n. Here $F_i(.)$ refers to the distribution function of the one-dimensional margin X_i.

Let $\mathbf{X} \in MAGH_n(0, \mathbf{I}, w)$. Then the corresponding copula coincides with independence copula, i.e. the copula density function is given by $c(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i$.

We finish this section with several comments regarding the advantages and disadvantages of the MGH and MAGH distributions. We start with the *distributional flexibility* to fit real data. An outstanding property of MAGH distributions is that, after an affine-linear transformation, all one-dimensional marginals can be fitted separately via different generalized hyperbolic distributions. In contrast to this, the one-dimensional marginals of MGH distributions are not that flexible since the parameters α and λ relate to the entire multivariate distribution and determine a strong structural behavior (see Definition 1). However, this structure causes a large subclass of MGH distributions to belong to the family of elliptically contoured distributions which inherit many useful statistical and analytical properties from multivariate normal distributions. For example, the family of elliptical contoured distribution is closed under linear regression and passing to the marginal distributions, see Cambinas at al. (1981).

Regarding the *dependence structure*, the MAGH distributions may have independent marginals for some parameter constellation. In particular, they support models which are based on linear combination of independent factors. In contrast, the MGH distributions are not capable to modelling independent marginals. They even yield extremal dependencies for bivariate distributions having correlation zero. Moreover, the correlation matrix of MAGH distributions is proportional to the scaling matrix Σ within large subclass of asymmetric MAGH distributions, whereas Σ is hardly to be interpret for skewed MGH distributions.

The copula of MAGH distributions, being the dependence structure of an affine–linearly transformed random vector with independent components, is quite illustrative and possesses many appealing properties. On the other hand, the copula structure of MGH distributions may suffer from inflexibility. Regarding the dependence of the extreme events, the MAGH distributions can model tail dependence whereas MGH distributions are always tail independent, as shown in Schmidt at al. (2006). Therefore, MAGH distributions are suitable especially within the field of risk management.

6 Conclusions

In this overview paper some alternative copula-based approaches to regression analysis are presented. For each one of the models considered a closed form expression for the joint distribution can be obtained, estimable by standard maximum likelihood techniques, and without the intermediate step of specifying the explicit distribution of unobserved factors that induce correlation. The copula approach produces dependence parameters that provide estimates of association between dependent variables. Convergence velocity is advantageous to copula approach too.

The copula methodology and the empirical model presented in Section 4 have many empirical applications. Any insurance in which two jointly determined outcomes are both directly or indirectly affected by the third variable could potentially be modelled with a similar formulation. The methodology proposed potentially may be applied for the following important problem from toxicology: the emission of two toxic pollutants from a factory might be related to profit expectations.

Nevertheless, the characterization of copulas as well as the choice of the dependence structure are difficult problems. For example, the choice of the copula does not inform explicitly which is the type of the dependence structure between variables involved, compare Examples 1 and 2 again.

As one can see, the primary task is just to choose an appropriate copula function, where the marginal distributions are treated as nuisance parameters. But what is the meaning of "appropriate"? In Fermanian and Scaillet (2005) are discussed some statistical pitfalls of copula modelling.

Copula-based regression analysis is profitable, if one knows the "right" copula, i.e. the "right" dependence structure, but if not - the procedure has to be repeated again, and again, until one find the "right" copula (which best fits the data according to some cost function).

Theoretically, the copula function is independent of marginals, and thus, copula is a very restrictive class of dependence functions. But, the geometrical behavior of the marginals (being increasing, decreasing, constant, unimodal functions, functions with a minimum, etc.), have their influence on the two-dimensional dependent structure, as demonstrated by Fernandez and Kolev (2007). The conclusions of this study just show that one should search for a new classes of dependent functions, in which the type of marginals should be taken into account. Goncalves and Kolev (2007) suggest to use the so-called "Sibuya dependence function" (where the marginal distributions have some impact of significance) as an alternative to the copula approach.

Acknowledgements

The authors acknowledge FAPESP (Grant 03/10105-2) and CAPES (Grant 171/04).

References

Ang, A., Chen, X. and Xing, X. (2002). Downside correlation and expected stock returns, Working paper No. 01-25, Columbia Business School. Available at http://ssrn.com/abstract=282986.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. Proceedings of the Royal Society London A, 401-419.

Benjamin, M., Rigby, R. Stasinopoulos, D. (2003). Generalized autoregressive moving average models. Journal of the American Statistical Association 98, 214-223.

Boyer, B.H., Gibson, M.S. and Loretan, M. (1999). Pitfalls in tests for changes in correlations, International Finance Discussion Paper No. 597, Board of Federal Reserve System.

Cambinas, S., Huang, S. and Simons, G. (1981). On the theory of the elliptically contoured distributions. Journal of Multivariate Analysis 11, 368-385.

Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. Journal of econometrics 135, 125-154.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). Copula Methods in Finance, Wiley Finance.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit model, Biometrika 85, 247-261.

Embrechts, P. McNeil, A. and Strauman, D. (2002). Correlation and dependence in risk management: properties and pitfalls, in: M. Dempster, H.K. Moffatt (Eds.), Risk Management: Value at Risk and Beyond, Cambridge Univ. Press, Cambridge, 176-223.

Engle, R.. (2002). Dynamic conditional correlation: A Simple class of multivariate GARCH models. Journal of Business and Economic Statistics 20, 339-350.

Engle, R. and Manganelli, S. (2007). CAViaR: Conditional autoregressive value at risk by regression quantiles. Forthcoming in Journal of Business and Economic Statistics.

Escarela, G., Mena, R. and Castill-Morales, A. (2006). A flexible class of parametric regression models based on copulas: application to poliomyelitis incidence. Statistical Methods in Medical Research 15, 593-609.

Fermanian, J-D. and Scaillet, O. (2005). Some statistical pitfalls of copula modelling for financial applications. In: Capital Formulation,Governance and banking, (E. Klein, Ed.), Nova Science Publishers, 59-74.

Fernandez, M. and Kolev, N. (2007). Bivariate density characterization by geometry of marginals. Forthcoming in Journal of Economic Quality Control.

Frees, J. and Wang, P. (2005). Credibility using copulas. North American Actuarial Journal 9, 31-48.

Frees, J. and Wang, P. (2006). Copula credibility for aggregate loss models. Insurance: Mathematics and Economics 38, 360-373.

Goncalves, M. and Kolev, N. (2007). Some probabilistic properties of Sibuya's dependence fiction. In: *Proc. Third Brazilian Conference on Statistical Modelling in Insurance and Finance*, (Eds. C. Fernandes and H. Schmidli), 336-339.

Joe, H. (1993). Parametric families of bivariate distributions with given marginals. Journal of Multivariate Analysis 46, 262-282.

Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman & Hall.

Kolev, N., Anjos, U. and Mendes, B. (2006). Copulas: a review and recent developments. Statistical Models 22, 617-660.

Magnus, W., Oberhettinger, F. and Soni, R. (1966). Formulas and Theorems for the Special functions of Mathematical Physics. Springer Verlag: Berlin.

Marshall, A. (1996). Copulas, marginals, and join distributions. In: *distributions with Fixed Marginals and Related Topics* (L. Ruschendorf, B. Sweitzer and M.D. Taylor (Eds.). Institute of Mathematical Statistics: Hayward, 213-222.

McNeil, A., Frey, R. and Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press.

McNeil, A. (2007). Sampling nested archimedian copulas. Forthcoming in Journal of Statistical Computation and Simulation.

Nelsen, R. (2006). An Introduction to Copulas, 2nd Edition, Springer: New York.

Niewiadomska-Bugaj, M. and Kowalczyk, T. (2005). On grade transformation and its implications for copulas. Brazilian Journal of Probability and Statistics 19, 125-137.

Pitt, M., Chan, D. and Kohn, R. (2006). Efficient Bayesian interface for Gaussian copula regression models. Biometrika 93, 537-554.

Song, P.X. (2000). Multivariate dispersion models generated from Gaussian copula. Scandinavian Journal of Statistics 24, 305-320.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris 8, 229-231.

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. Math. Methods in Operations Research 55, 301-327.

Schmidt, R., Hrycej, T. and Stütze, E. (2006). Multivariate distribution models with generalized hyperbolic marginals. Computational Statistics and Data Analysis 50, 2065-2096.

Szczesny, W. (1991). On the performance of a discriminant function, Journal of Classification 8, 201-215.

Tiit, E. and Kaarik, E. (1996). Generation and investigation of multivariate distributions having fixed discrete marginals. In: *distributions with Fixed Marginals and Related Topics* (L. Ruschendorf, B. Sweitzer and M.D. Taylor). Institute of Mathematical Statistics: Hayward, 113-127.

Train, K. (2003). Discrete Choice Models with Simulation. New York: Cambridge University Press.

Trivedi, P.K. and Zimmer, D.M. (2006). Copula Modelling in Econometrics: Introduction to Practitioners. Wiley: New York.

Van Ophem, H. (1999). A general method to estimate correlated discrete random vector. Econometric Theory 15, 228-237.

Zimmer, D. and Trivedi, P. (2006). Using trivariate copulas to model sample selection and treatment effect: application to family health care demand. Journal of Business and Economic Statistics 24, 63-76.

Appendix 1: Particular Cases of Elliptical Distributions

Let $\mathbf{Y} \in E_n(\mu, \Sigma, g)$ and C_n be a normalizing constant.

- **Multivariate spherical distributions.** The random vector $\mathbf{X}$ in Definition 1 is *spherically distributed* if the density generator $g(.)$ in (19) is given by

$$g(u) = C_n \frac{K_{\lambda - n/2}(\alpha\sqrt{1+u})}{(1+u)^{n/4-\lambda/2}}, \quad u \in \Re;$$

- **Multivariate normal and Kotz type of distributions.** The random vector $\mathbf{Y}$ has *symmetric Kotz type distribution* if the density generator $g(.)$ in (19) is of the form

$$g(u) = C_n u^{m-1} \exp(-ru^s), \quad r, s > 0, \ 2m + n > 2.$$

  The *multivariate normal distribution* is a particular case if $m = s = 1$ and $r = 0, 5$;

- **Multivariate t-distributions and symmetric Pearson type VII distributions.** The random vector $\mathbf{Y}$ has *symmetric Pearson type VII distribution* if the density generator $g(.)$ in (19) is of the form

$$g(u) = C_n \left(1 + \frac{u}{s}\right)^{-m}, \quad s > 0, \ 2m > n.$$

  Setting $m = (n+k)/2$ yields the density generator of *multivariate t-distributions*, which include the *multivariate Cauchy distribution* for $k = 1$.

- **Multivariate logistic distributions.** The random vector has *multivariate logistic distribution* if the density generator $g(.)$ in (19) is of the form

$$g(u) = C_n \frac{\exp(-u)}{[1 + \exp(-u)]^2}.$$

- **Multivariate symmetric generalized hyperbolic distributions.** The random vector $\mathbf{Y}$ has *multivariate symmetric generalized hyperbolic distribution* if the density generator $g(.)$ in (19) is of the form

$$g(u) = C_n \frac{K_{\lambda-n/2}[\sqrt{\psi(\chi+u)}]}{(\sqrt{\chi+u})^{n/2-\lambda}}.$$

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2007-01 - POLPO, A., PEREIRA, C.A.B. Reliability nonparametric bayesian estimation in parallel systems 2007. 12p. (RT-MAE 2007-01)

2007-02 - GONÇALVES, M., Kolev, N. V., Fabris. A. E. Bounds for quantile-based measures of dependent risks' functions. 2007. 16p. (RT-MAE 2007-02)

2007-03 - MIMBELA, J. A. L, Kolev, N.V. Occupation measure of Markov - modulated risk processes 2007. 12p. (RT-MAE 2007-03)

2007-04 - GONÇALVES, M., Kolev, N. V., Fabris, A. E. Bounds for distorted risk measuress. 2007. 11p. (RT-MAE 2007-04)

2007-05 - BUENO, V.C., A series representation of a parallei systems. 2007. 10p. (RT-MAE 2007-05)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasi