Copula-Based Regression Models

Nikolai Kolev[1] **and Delhi Paiva**[2]

[1]Department of Statistics,
University of São Paulo
[2]EACH, University of São Paulo
nkolev@ime.usp.br and delhi@usp.br

Tartu, June 26, 2007

Outline

1. Copulas (1)

Let $F(y_1, y_2)$ be the joint distribution function of random variables Y_1 and Y_2 whose distributions $F_1(y_1) = P(Y_1 \leq y_1)$ and $F_2(y_2) = P(Y_2 \leq y_2)$ are **continuous.**

Denote the inverse functions by $F_i^{-1}(v_i) = inf\{y_i : F_i(y_i) \geq v_i\}$, for $i = 1, 2$.

Sklar's theorem, e.g., Sklar (1959), states that **there exists a unique** copula function $C(.,.)$ such that

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

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

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

1. Copulas (2)

Consequently

- **Copulas allow** one **to model the marginal distributions** and the **dependence structure** of multivariate random variable **separately.**
- Furthermore, the **functional association** between underlying random variables **is not influenced by the marginal behavior.**

Example 1

Let Y_1 and Y_2 be r.v.'s with joint distribution

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

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 are r.v.'s with joint distribution

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

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 inverses are

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

The related copula is $C(v_1, v_2) = \frac{v_1 v_2}{v_1+v_2-v_1 v_2}$, i.e. the **same** as in Example 1. **Is that fact surprising ?**

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 (that **captures the degree of associaton** between the univariate marginals) 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

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

1. Gaussian copula (6)

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 defined by

$$C(v_1, v_2; \rho) = \Phi_2(\Phi^{-1}(v_1), \Phi^{-1}(v_2); \rho), \quad v_1, v_2 \in (0, 1), \qquad (2)$$

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 joint distribution

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

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 just the Gaussian copula (2).

1. Copulas - discrete case (7)

It is possible to derive a **unique** copula representation 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 $P(Y = y_r) = p_r,\ r \geq 1$. Then, $P(F_Y(Y) \leq u) \leq u$, for $u \in (0,1]$ and **the distribution** of $F_Y(Y)$ **is discrete one**, given by

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

and therefore, **faraway from the uniform distribution on (0,1)**. The problem is, that in the discrete case $F_Y(Y)$ is one-to-many, while $F_Y^{-1}(Y)$ is many-to-one functions.

1. Copulas - discrete case (8)

Let us underline that a version of Sklar's Theorem is offered recently by Niewiadomska-Bugaj and Kowalczyk (2005) who find a **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 a theoretical issue, but it **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.

1. Copulas - discrete case (9)

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.

Regression estimation.

In discrete case, the copula density $c(.,.)$ can be formed by taking differences

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

for the corresponding copula probability mass function.

The related log-likelihood function is maximized using a quasi-Newton iterative algorithm requiring only first derivative.

2. The Gaussian copula regression model (1)

Suppose there are n observations $\mathbf{y_1}, ..., \mathbf{y_n}$, each of dimension p. The **Gaussian copula regression model** can be defined as

$$y_{ij} = h_{ij}^{-1}(z_{ij}), \quad j = 1, \ldots, p, \quad \mathbf{z}_i \sim N_p(0, \mathbf{R}) \tag{3}$$

for $i = 1, \ldots, p$, see Pitt et al. (2006). In (3), $h_{ij}^{-1}(.) = F_{ij}^{-1}(\Phi(.))$, where $F_{ij}(.)$ 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 of the vector parameter θ_j and $m \times 1$ vector of covariates x_{ij}.

2. The Gaussian copula regression model (2)

Typically we might have a marginal generalized linear model for a given y_{ij}.

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 in the model associated with the j-th component.

This approach assumes that the **marginal distributions 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**.

2 The Gaussian copula regression model (3)

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 Gaussian copula regression model with identical marginals whose parameters are known, providing a method for estimating the parameters of discrete marginals.

2. The Gaussian copula regression model: Example 3 (4)

Example 3: 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**. One may 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.

2. The Gaussian copula regression model: Example 3 (5)

The error terms are standardized to have a variance 1:

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

where t_{ν_j} is a t-distribution with $\nu_j > 2$ degrees of freedom. 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 $\mu_{ij} = x'_{ij}\beta_j$.

Remark 3: Difference between discrete and continuous models.

Although the Gaussian copula regression model

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

with $h_{ij}^{-1}(.) = F_{ij}^{-1}(.)$, $F_{ij}(.) = F_j(.; \theta_j; x_{ij})$, $i = 1, \ldots, n$ 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), see Pitt et al. (2006).

3. Transition regression models based on copulas (1)

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 themself 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 (1)

To establish context, we briefly review the first order regression model for normal responses. Consider the stationary time-series $\{Y_t, t = 1, 2, \ldots\}$ with marginal responses $Y_t \sim N(\beta^{\mathbf{T}}\mathbf{x}_t, \sigma^2)$ for $t = 1, 2, \ldots$.

Thus,

- $\beta^{\mathbf{T}}\mathbf{x}_t$ 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.

3.1. Construction of first order copula transition models (2)

If the correlation between lagged responses Y_{t-1} and Y_t is $\rho \in (-1, 1)$, the transition model has the following specification

$$Y_t | Y_{t-1} \sim N(\beta^{\mathbf{T}} \mathbf{x}_t + \rho[Y_{t-1} - \beta^{\mathbf{T}} \mathbf{x}_{t-1}]; \sigma^2(1 - \rho^2)). \qquad (4)$$

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 $F_{2|1}(y_2|y_1)$ can be computed as

$$F_{2|1}(y_t|y_{t-1}) = P(Y_t \leq y_t | Y_{t-1} = y_{t-1}) = \frac{\partial F(y_{t-1}, y_t)/\partial y_{t-1}}{\partial H(y_{t-1})/\partial y_{t-1}}.$$

3.1. Construction of first order copula transition models (3)

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

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

the corresponding conditional density function $f_{2|1}(y_2|y_1)$ can be written using the copula density function as

$$f_{2|1}(y_2|y_1) = h(y_2) c(H(y_1), H(y_2)). \tag{5}$$

3.1. Construction of first order copula transition models (4)

If one wishes to obtain a first-order Gaussian autoregressive model in terms of copula transition model corresponding to

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

he simply needs to substitute in

$$f_{2|1}(y_t | y_{t-1}) = h(y_t) c(H(y_{t-1}), H(y_t)),$$

the Gaussian copula density function $c(., .; \rho)$ from

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

the standardized marginal distribution $H(y_t) = \Phi[(y_t - \beta^{\mathbf{T}} \mathbf{x}_t)/\sigma]$ and its density $h(y_t)$.

3.1. Construction of first order copula transition models: Remark (5)

Remark: 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:

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

3.1. Construction of first order copula transition models: Remark 4 (6)

The corresponding transition density function is

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

where the probability mass function $c(y_t, y_{t-1})$ should be calculated using differences

$$\begin{aligned} c(y_t, y_{t-1}) = {} & C(F_1(y_{t-1}), F_2(y_t)) - C(F_1(y_{t-1} - 1), F_2(y_t)) \\ & - C(F_1(y_{t-1}), F_2(y_t - 1)) + C(F_1(y_{t-1} - 1), F_2(y_t - 1)). \end{aligned}$$

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).

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 (CVaR) 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 function**, remind

$$f_{2|1}(y_t|y_{t-1};\theta) = h(y_t)c(H(y_{t-1}), H(y_t);\theta).$$

3.2. Semiparametric copula regression models (2)

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 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

$$f_{2|1}(y_t|y_{t-1};\theta) = h(y_t)c(H(y_{t-1}), H(y_t);\theta).$$

we can write the conditional density of Y_t given Y_{t-1} by

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

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 using the copula approach by letting copula parameter θ depending on x_{1t}, say, and the marginal distribution $G^*(.)$ depending on x_{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.

3.2. Semiparametric copula regression models: Observation (5)

Observation: The transformed process, $\{U_t : U_t = G^*(Y_t)\}$, is a **stationary parametric Markov process**.

Since 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: Semiparametric model generated by the Gaussian copula.

Let the copula $C(.,.;\rho)$ be the Gaussian copula

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

Then the process $\{\Phi^{-1}(G^*(Y_t))\}$ is a Gaussian process that can be represented by

$$\Phi^{-1}(G^*(Y_t)) = \rho\Phi^{-1}(G^*(Y_{t-1})) + \epsilon_t, \tag{6}$$

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 (6):

- if $G^*(.)$ is the **standard normal distribution**, then $\{Y_t\}$ is a linear AR(1) process;
- if $G^*(.)$ **is unspecified**, then we have the class of semiparametric model generated by the Gaussian copula;
- if $G^*(.)$ **is 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

$$\Lambda_{1,\theta_1}(G^*(Y_t)) = \Lambda_{2,\theta_2}(G^*(Y_{t-1})) + \sigma_{\theta_3}(G^*(Y_{t-1}))e_t, \qquad (7)$$

where

- $G^*(.)$ 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 $\{e_t\}$ are 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 (7) 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

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

$\Lambda_{1,\theta_1}(u_1) = \Phi^{-1}(u_1)$, $\Lambda_{2,\theta_2}(u_0) = \rho\Phi^{-1}(u_0)$, $\sigma_{\theta_3}(u_0) = \sqrt{1-\rho^2}$, $h(., \theta_4)$ 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.

4. Copula-based longitudinal models (1)

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 claim amount** S_{it} and **claim 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

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

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

4. Copula-based longitudinal models (2)

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}), 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**.

4. Copula-based longitudinal models (3)

Consider the history of the i-th risk class and let $H^N_{i,t-1} = \{N_{i,1}, \dots, N_{i,t-1}\}$ be **claims number history** $H_{i,t-1} = \{S_{i,1}, N_{i,1}, \dots, S_{i,t-1}, N_{i,t-1}\}$ and the **claim 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 claim process, i.e. the **distribution of S_{it} is a function of N_{it} and $H_{i,t-1}$**;
- Claim number depends on the prior history of the claims number process but not the claim severity process. That is, **the distribution of N_{it} is a function of $H^N_{i,t-1}$.**

Under these assumptions, the joint distribution $F(.)$ of $\{S_{i,1}, N_{i,1}, \dots, S_{i,T}, N_{i,T}\}$ may be written as

$$F(S_{i,1}, N_{i,1}, \dots, S_{i,T}, N_{i,T}) = \{\text{frequency distribution}\} \times \{\text{conditional severity distribution}\}$$

4. Copula-based longitudinal models (4)

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 claim severity process one can use:

- **a generalized linear model for the marginal distributions** (to describe the cross-sectional characteristics), **conditional of frequency**;
- **a parametric copula to model the joint distribution of claims over time.**

4. Copula-based longitudinal models (5)

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 and their joint distribution is represented via $t-$ and Gaussian copulas. The authors 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, and are described in Frees and Wang (2005).

[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.

4.2. MGH distributions and their copulas (1)

Since copulas are concerned primarily with relationships, one **can use any multivariate distribution to generate a copula**. Here we propose the use of parametric copula corresponding to the multivariate distribution with **generalized hyperbolic margins in a longitudinal data framework**.

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 **suffers from not having 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

$\mathbf{Y} \overset{d}{=} \mathbf{A}'\mathbf{X} + \mu$ for some lower triangular matrix $\mathbf{A}' \in \Re^{n \times n}$ such that $\mathbf{A}'\mathbf{A} = \Sigma$ is positive definite and $\mathbf{X}$ has a density function of the form ($\mathbf{x} \in \Re^n$):

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{C_n K_{\lambda - n/2}(\alpha\sqrt{1 + \mathbf{x}'\mathbf{x}})}{(1 + \mathbf{x}'\mathbf{x})^{n/4 - \lambda/2}} \exp(\alpha\beta'\mathbf{x}) \quad \text{with} \quad C_n = \frac{\alpha^{n/2}(1 - \beta\beta')^{\lambda/2}}{(2\pi)^{n/2} K_\lambda(\alpha\sqrt{1 - \beta\beta'})}. \tag{8}$$

$K_\nu(.)$ 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 $\| \beta \|_2 < 1$ ($\| \, . \, \|_2$ 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 (8) **is its invariance under affine-linear transformations.**

For

$\lambda = \frac{n+1}{2}$ we obtain **multivariate hyperbolic density**

$\lambda = -1/2$ the **multivariate inverse Gaussian** density.

$\lambda = 1$ leads to **hyperbolically distributed one-dimensional marginals.**

4.2. MGH distributions and their copulas (4)

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

$$f_{\mathbf{Y}}(\mathbf{y}) = |\Sigma|^{-1/2} g[(\mathbf{y} - \mu)' |\Sigma|^{-1} (\mathbf{y} - \mu)], \quad \mathbf{y} \in \Re^n, \tag{9}$$

for some density *generator function* $g(.) : (0, \infty) \to (0, \infty)$. Let us denote the family of n-dimensional **elliptically contoured distributions**[2] by $E_n(\mu, \Sigma, g)$ For a detailed treatment of elliptically countered distributions see Schmidt (2002, 2006). Particular cases of (9) are listed in Appendix 1.

[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 is independent of the m-dimensional random vector $\mathbf{U}^{(m)}$ is uniformly distributed on the unit sphere in $\Re^m$.

5. A copula based regression model for unobserved heterogeneity (1)

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 related**, i.e. if a husband's and wife's utilization and their insurance choice are simultaneously determined.

5.1. Regression specifications of the model (1)

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

$$
\begin{aligned}
Y^U_{i,w} &= \mathbf{x^T}_{i,w}\beta_w + u_{i,w} + \lambda_{v,w}v_i, && (10)\\
Y^U_{i,h} &= \mathbf{x^T}_{i,h}\beta_h + u_{i,h} + \lambda_{v,h}v_i && (11)\\
D^U_i &= \mathbf{z^T}_i\alpha + \lambda_u(u_{i,w} + u_{i,h}) + \epsilon_i, && (12)
\end{aligned}
$$

- λ_u measure how the insurance choice is related to family consumption;
- $\lambda_{v,w}$ and $\lambda_{v,h}$ indicate the degree of wifes and husbands contribution to utlization;
- $x_{i,w}$ and $x_{i,h}$ 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;
- ϵ_i are independently distributed error terms, assumed to be uncorrelated with $(u_{i,w}, u_{i,h}, v_i)$;
- quantities u_i and v_i represent the **dependence between spouses' utilization and insurance choice** and are **not observed** by the researcher.

5.1. Regression specifications of the model (2)

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 (and husband's) utilization probability for a particular medical service is assumed to follow a **negative binomial (NB)** distribution in the form

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

where $\mu_{i,w} = exp(\mathbf{x^T}_{i,w}\beta_w)$ is the **conditional mean**.
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

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

5.2. Trivariate Frank copula specification (1)

The copula's facility to accommodate unobserved heterogeneity is especially important for the data analyzed. **Joint estimation to single equation estimation**, (used in regression analysis) is profitable 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 $F_2(.)$ of the copula selected are wife's NB and husband's NB distributions

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

and $f_2(y_{i,h}|\mu_{i,h})$ respectively. The third marginal, $F_3(.)$ of the corresponding copula is the Probit model

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

5.2. Trivariate Frank copula specification (2)

Consider the trivariate mixtures of powers representation

$$C(u, v, w, \theta_1, \theta_2) = \int_0^\infty \int_0^\infty B(u|\theta_2)B(v|\theta_2)dM_2(\theta_2; \theta_1)B(w|\theta_1)dM_1(\theta_1) \quad (13)$$

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 (11) **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_1 measures the **dependence between u and w** (θ_{12} say) **as well as between v and w**, (θ_{23} say) and the two **must** be equal (i.e. $\theta_1 = \theta_{12} = \theta_{23}$).

Thus, the copula representation (13) **accommodates unobserved heterogeneity in the tree outcomes.**

5.2. Trivariate Frank copula specification (3)

As a particular case we obtain the **trivariate Frank copula**

$$C_F(u, v, w, \theta_1, \theta_2) = -\theta_1 log \left\{ 1 - \left(1 - [1 - c_2^{-1}(1 - e^{-\theta_2 u})(1 - e^{-\theta_2 v})]^{\theta_1/\theta_2} \right) \frac{1 - e^{-\theta_1 w}}{c_1} \right\} \tag{14}$$

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 (14) 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 (13)), e.g. Zimmer and Trivedi (2006).

In

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

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, 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 $f(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, u_w, u_h, v)$ of the trivariate Frank model

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

can be decomposed by Bayes' rule as follows

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

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

$$f_{12|3}(y_w, y_h|\mathbf{x}_w, \mathbf{x}_h, D, v) = \frac{f(y_w, y_h, D|\mathbf{x}_w, \mathbf{x}_h, \mathbf{z}, u_w, u_h, v)}{f_3(D|\mathbf{z})} \quad (15)$$

For the data considered and for each of utilization measures (physician visits (PV), nonphysician visits (NPV) and emergency room visits (ER), as in regression analysis of the insurance data), **the numerator** in

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

is estimated by Frank copula

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

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. *Comparision*

	Physician visits		Non physician		Emergency	
	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	
Regression Model						
Selection Coefficient (λ_u)	0.023	0.024	-0.052	0.042	0.035	0.056
Coefficient ($\lambda_{v,h}$)	0.438	0.034	0.885	0.082	0.766	0.154
Log Likelihood	-30743.33		-16969.16		-6918.72	

Observation: The calculated values of the correponding log-likelihood functions for the both models are different, because are based on the different compositions.

5.2. Trivariate Frank copula specification (8)

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 correlation between spouses' utilization** (measured by θ_2 in the copula model and $\lambda_{v,h}$ in the regression specificationl)
- correlation is largest in magnitude for nonphysician utilization (NPV) followed by ER (emergency visits) physician usage (PV).

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**.

6. Conclusions (1)

In this overview, some alternative copula-based approaches to regression analysis are presented, with many potential applications. 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.**

6. Conclusions (2)

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 have to be repeated again, and again, until one find the "appropriate"copula.

6. Conclusions (4)

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 **marginal densities** (being increasing, decreasing, constant, unimodal functions, functions with a minimum, etc.), **have their influence on the two-dimensional dependent structure**, as demonstrated by Fernadez 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**.

A suggestion is made in Kolev, Gonçalves and Dimitrov (2007) to use the so-called "Sibuya's dependence function".

References (1)

- 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.

References (2)

- Cherubini, U., Luciano, E. and Vecchiato, V. (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.

References (3)

- 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.
- Fernadez, M. and Kolev, N. (2007). Bivariate density characterization by geometry of marginals. Forthcoming in Journal of Economic Quality Control 22.
- 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.
- Gonçalves, M. and Kolev, N. (2007). Some probabilistic properties of Sibuyat's dependence function. Proceedings of the Third Brazilian Conference on Statistical Modelling in Insurance and Finance.
- 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.

References (4)

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

References (5)

- 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 margins. Computational Statistics and Data Analysis 50, 2065-2096.

References (6)

- 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.