The Step Construction of Copula Gaussian Multivariate and AR(1)-N.GARCH(1,1) Models

Rezzy Eko Caraka[12*], Hasbi Yasin[3**] , Toni Toharudin [2***]
{rezzyekocaraka@gmail.com[*], hasbiyasin17@gmail.com[**] , toni.toharudin@unpad.ac.id [***] }

School of Mathematical Sciences, The National University of Malaysia[1],
Department of Statistics Padjadjaran University, Bandung, Indonesia [2],
Department of Statistics, Diponegoro University, Indonesia [3]

Abstract. Copulas are a powerful tool in multivariate statistics. If copula functions used for modeling dependence between random variables, there is an immediate and obvious need to test whether the model can describe the data at hand accurately enough or not. Copulas involve several underlying functions: the marginal cumulative distribution functions (CDF) and a joint CDF. To estimating the copula functions, first issue consists in specifying how to estimate the margins separately. Moreover, some of these functions can be fully known. Depending on the assumptions made, some quantities have to be estimated parametric, semiparametric or even non-parametric. In this paper, we propose the step of parameter estimation in modeling Copula Gaussian Multivariate and AR (1)-N. GARCH (1,1) Models

Keywords: Copula, Multivariate, GARCH, Time Series

1 Introduction

The copula is one of the statistical methods that can describe the relationship between variables that are not too strict on the assumption of distribution and can show the relationship of dependencies on extreme points. Mostly copula used in finance, climate, environmental (Kazianka, 2013). This method can describe the dependency structure between variables with different margins and model their tail dependencies. Many copula models can be used such as Copula Gaussian (Kolev & Paiva, 2009), Copula-t, Copula Clayton (Patton, 2013), Copula Frank, and Copula Gumbel(Choroś, Ibragimov, & Permiakova, 2010) . In this paper will be demonstrated the application and the step construction of Copula gaussian with AR (1)-N. GARCH (1,1). A time series $\{Y_t, t \in T\}$ is said to be stationary if f every t is applicable:

a. $E[Y_t] = \mu$ (constant)
b. $cov(Y_t . Y_{t-k}) = \gamma_k$ (not dependent on t)

The stationary data TS $\{Y_t, t \in T\}$ fluctuates around its mean with constant variance. The method to stationary the data one of them is log return.

$$Y(t+1) = \log\left(\frac{S(t+1)}{S(t)}\right)$$

If $S(t+1)$ denotes the stock price at time $t+1$, , then the log return of the stock price at time $t+1$, expressed by $Y(t+1)$. At $t = 0$, it is assumed $Y(0) = 0$ to keep much data the same.

2 The Step Construction

Here is the formula for model AR (1) - N. GARCH (1,1)

$$Y_t = \mu + \alpha Y_{t-1} + \sigma_t \epsilon_t,\ \epsilon_t \sim N(0,1) \tag{1}$$

$$\sigma_t^2 = \alpha_0 + \beta_1 \left(Y_{t-1} - \mu - \alpha Y_{(t-2)}\right)^2 + \beta_2 \sigma_{t-1}^2 \tag{2}$$

with parameters $\alpha_0, \beta_1, \beta_2$.

Here are the steps to estimate the parameters of AR (1)-N. GARCH (1,1) models:

1. Assume the error value ϵ_tt is normally distributed (0.1).
2. Determine the parameter values $\alpha_0, \beta_1, \beta_2$ by using the Likelihood function defined as follows:

$$P\left(\sigma_{i,t}\epsilon_{i,t} \leq s_t \,|\Omega_{t-1}\right)$$

$$F(s_t \,|\Omega_{t-1}) = P\left(\epsilon_{i,t} \leq \frac{s_t}{\sigma_{i,t}} \,\middle|\Omega_{t-1}\right)$$

$$= \int_{-\infty}^{\frac{s_t}{\sigma_{i,t}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$\text{pdf } \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s_t}{\sigma_{i,t}}\right)^2}$$

$$f(s_t) = \left(\sigma_{i,t}\epsilon_{i,t} \,|\Omega_{t-1}\right) = \left(\theta_{i,t} \,|\Omega_{t-1}\right) \text{ with } \frac{1}{\sigma_{i,t}} \sim N\left(0, \sigma_{i,t}^2\right)$$

$$\text{Then } P\left(\theta_{i,2}, \theta_{i,3}, \dots, \theta_{i,T}\right) =$$

$$f\left(\theta_{i,t} \,|\theta_{i,T-1}, \theta_{i,T-2}, \dots, \theta_{i,2}\right). f\left(\theta_{i,T-1} \,|\, \theta_{i,T-2}, \theta_{i,T-3} \dots, \theta_{i,2}\right) \dots f\left(\theta_{i,3} \,|\, \theta_{i,2}\right) f(\theta_{i,2})$$

$$f\left(\theta_{i,2}, \theta_{i,3}, \dots, \theta_{i,T}\right) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s_t}{\sigma_{i,t}}\right)^2}.$$

Select $\sigma_{i,1} = 0$ or the standard deviation value of the data.

The Likelihood log function is as follows:

$$\log f = \log \frac{1}{\sqrt{2\pi}} + \left(\sum_{t=2}^{T} -\frac{1}{2}\left(\frac{s_t}{\sigma_t}\right)^2\right) \tag{3}$$

To obtain the parameters, maximize the likelihood log function above by using the optimization method.

3. Enter the values of parameters that have been obtained from AR model (1) - N.GARCH (1,1) to get error value ϵ_t from the model.
4. Testing the error of AR (1)-N.GARCH (1,1) using ACF and PACF plots.
5. Perform a distribution fitting to find out the distribution of error data.

The copula is a joint distribution function of some marginal distribution function(Genest & Favre, 2007). The copula is used to analyse the dependence of random variables in structures described by the combined function (Caraka, Yasin, Sugiyarto, Sugiarto, & Ismail, 2016) and (Caraka, Supari, & Tahmid, 2018). Here is Gaussian copula for two variables, which can be easily extended to multivariable.

Let $(X, Y) \sim$ Bivariate Normal $X \sim N(0,1), Y \sim N(0,1)$

h (x, y) = f (x) .g (y) .c (u, v), where $u = F(x)$, $v = G(y)$

Pdf Gaussian Copula

$$C(u, v) = \frac{1}{|\Sigma|^{1/2}} e^{1/2([x\ y](I - \Sigma^{-1})\begin{bmatrix} x \\ y \end{bmatrix})}$$

When $x = \Phi^{-1}$(u), y=Φ^{-1}(v), with $\Phi(x)$ is the cumulative and normal standard distribution.

At the same time, gaussian copula can describe as equation (4):

$$C_\Lambda(u_1,\ldots,u_T) = \Phi_{0,\Lambda}(\Phi_{0,1}{}^{-1}(u_1),\Phi_{0,1}{}^{-1}(u_2),\ldots,\Phi_{0,1}{}^{-1}(u_T))$$

$$\Phi_{0,\Lambda}(x_1,\ldots,x_T)=\int_{-\infty}^{x_T}\int_{-\infty}^{x_{T-1}}\ldots\int_{-\infty}^{x_1}\frac{1}{(2\pi)^{T/2}\,|\Lambda|^{1/2}}\,e^{-\frac{1}{2}\left[(s_1,\ldots,s_T)\Lambda^{-1}\begin{pmatrix}s_1\\ \vdots\\ s_T\end{pmatrix}\right]}$$

$$\Phi_{0,1}(\mathrm{x}) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}s^2}\,ds \quad , \; u_i = F_i(x_i)$$

$$C_\Lambda(u_1,\ldots,u_T) = \Phi_\Lambda(\Phi_{0,1}{}^{-1}F_1(x_1),\Phi_{0,1}{}^{-1}F_2(x_2),\ldots,\Phi_{0,1}{}^{-1}F_T(x_T))$$

$$\text{if } F_i(.)\sim\Phi_{\mu_i,\sigma_i{}^2}(.), \text{then}$$

$$C_\Lambda(u_1,\ldots,u_T) \;=\; \Phi_\Lambda(\Phi_{0,1}{}^{-1}(\Phi_{\mu_{1,\sigma_1{}^2}}(x_1)),\Phi_{0,1}{}^{-1}(\Phi_{\mu_{2,\sigma_2{}^2}}(x_2)),\ldots,\Phi_{0,1}{}^{-1}(\Phi_{\mu_{T,\sigma_T{}^2}}(x_T)))$$

$$= \;\Phi_\Lambda\left(\Phi_{0,1}{}^{-1}\left(\Phi_{0,1}\left(\frac{x_1-N_1}{\sigma_1}\right)\right),\Phi_{0,1}{}^{-1}\left(\Phi_{0,1}\left(\frac{x_2-N_2}{\sigma_2}\right)\right),\ldots,\Phi_{0,1}{}^{-1}\left(\Phi_{0,1}\left(\frac{x_T-N_T}{\sigma_T}\right)\right)\right) \quad (4)$$

Let $X_i\sim F$ copula gauss is defined as

$$C_\Lambda = \Phi_\Lambda\big(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_T)\big), \quad (5)$$

With:

Φ_Λ States the function of a cumulative distribution a multivariate normal with a covariance matrix Λ,

Φ^{-1} inverse normal standard univariate CDF, and $u_i = F(x_i)$

Probability density function (pdf) from $F(x)$ is $f_X(x)$,can be defined as:

$$f_X(x) = \frac{1}{(2\pi)^{\frac{T}{2}}|\sum|^{\frac{1}{2}}}e^{-\frac{1}{2}(x\;\mu)^T\Sigma^{-1}(x\;\;\mu)} \quad (6)$$

with

$$x' = (x_1, x_2, \ldots, x_T)$$

$$\sum \;=\begin{bmatrix}\mathrm{var}(\mathrm{x}_1,\mathrm{x}_1) & \mathrm{cov}(\mathrm{x}_1,\mathrm{x}_2) & \cdots & \mathrm{cov}(\mathrm{x}_1,\mathrm{x}_\mathrm{T})\\ \vdots & \ddots & \vdots & \vdots\\ \mathrm{cov}(\mathrm{x}_\mathrm{T},\mathrm{x}_1) & \cdots & \cdots & \mathrm{var}(\mathrm{x}_\mathrm{T},\mathrm{x}_\mathrm{T})\end{bmatrix}$$

Then:

$$F_X(y_1, y_2, \ldots, y_T) = \int_{-\infty}^{y_T}\int_{-\infty}^{t_{T-1}}\ldots\int_{-\infty}^{y_1}f(x)\,dx_1dx_2\ldots dx_T$$

$$C_\Lambda = \int_{-\infty}^{\Phi^{-1}(u_T)}\ldots\int_{-\infty}^{\Phi^{-1}(u_1)}f(x)\,dx_1\ldots dx_T \quad (7)$$

Since the correlation between variables is unknown, the correlation is approximated by utilizing the profile information of the data, which in this case is a partition time index. The variance-covariance matrix is approximated by the Kernel matrix, which has a similar trait, namely positive and symmetric definite. The Kernel matrix is chosen to maximize the likelihood function value of Copula Gauss. The kernel functions used are:

$$\Lambda_T(p,q) = \;h_j\,e^{\frac{|x_p-x_q|^2}{h_j}} \quad (8)$$

From the previously obtained Copula Gauss, we can obtain the loglikelihood function to obtain kernel parameters as an approximation of the variance-covariance matrix.

$$log\;C_{\Lambda_T}(u_1,\ldots,u_T) = \frac{1}{2}\log|\Lambda_T| + \frac{1}{2}\,z^T(I-\Lambda_T^{-1})\,z \quad (9)$$

Next, we will find the parameter estimate h that maximizes likelihood function with the following algorithm.

Algorithm

1. Choose h_j

2. Get the kernel

$$\Lambda_T(p,q) = h_j\, e^{\frac{|x_p - x_q|^2}{h_j}} \quad with \quad p,q = 1,2,\dots,T$$

3. Determine the lower triangular matrix L by Cholesky factorization on the Λ_T matrix already obtained above

$$\Lambda_T = LL'$$

4. Complete the equation $Lv = z$.
   with $z_j = \phi^{-1}(u_j)$, $u_j = F(x_j)$ in this case x_j use the data of $\epsilon_{i,j}$
   Calculate $L_j = -\sum_{j=1}^{T} \log L_{jj} + \frac{1}{2} z'z - \frac{1}{2} v'v$
5. Select h with condition has the largest L_j.

However, , we propose Copula binding (Ma & Sun, 2011) to combine two or more copulas. Especially for Copula Gauss Bivariate, pdf from copula can be written as:

$$c(u,v:\varrho) := \frac{\partial^2}{\partial u \partial v} C(u,v:\varrho) = \frac{\varphi_2(\Phi^{-1}(v);\Phi^{-1}(v);\varrho)}{\varphi(\Phi^{-1}(u);\varphi^{-1}(\Phi^{-1});v)}$$
$$= \frac{1}{\sqrt{1-\varrho^2}} \exp\left(\frac{2\varrho\Phi^{-1}(u)\Phi^{-1}(v) - \varrho^2(\Phi^{-1}(u)^2\Phi^{-1}(v)^2)}{2(1-\varrho^2)}\right. \quad (10$$

With CDF as follows:

$C(u,v:\varrho) = \int_0^u \int_0^v c(s,t;\varrho)dt\, ds$

In this case, ρ is a measure of Pearson's dependence. For elliptic copulas such as Gaussian Copula, ρ can be obtained as follows:

$$\rho = \sin\left(\frac{\pi}{2}\tau\right) \quad (11)$$

With τ representing coefficient of correlation Kendall's Tau. Suppose $(x_1, y_1), \dots (x_n, y_n)$ the set of random variables of X and Y respectively, so that each pair of values (x_i, y_i) is single. Any observation pair (x_i, y_i) and (x_j, y_j) are defined to be concordant if the order of each element is equal, in other words $x_i < x_j$ and $y_i < y_j$ or $x_i > x_j$ and $y_i > y_j$. Kendall's Tau is calculated as follows:

$$\tau = \frac{\#Concordant\ correlation - \#\ \text{Disconcordant correlation}}{\frac{1}{2}(n)(n-1)} \quad (12)$$

3. PREDICTION

In this section, we will provide a predictive explanation of predictions using copula(Patton, 2013). The formula of the conditional distribution function of AR-N.GARCH model error for $F(\varepsilon_2|\varepsilon_1)$.

$$F(\varepsilon_2|\varepsilon_1) = \frac{\frac{\partial}{\partial u_1} C(u_1,u_2)}{c(u_1)=1}$$
$$= \int_{-\infty}^{\Phi^{-1}(u_2)} f(\Phi^{-1}(u_1), s_2) ds_2 \frac{\partial \Phi^{-1}(u_1)}{\partial u_1}$$
$$= \int_{-\infty}^{\Phi^{-1}(u_2)} f(\Phi^{-1}(u_1), s_2) ds_2 \frac{1}{\phi(\tilde{\Phi}(u_1))} \quad \dots(*)$$

Let $y_1 = \Phi^{-1}(u_1)$, then the equation (*) becomes:

$$\int_{-\infty}^{y_1} \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_1 \ \ x_2)\Sigma^{-1}\binom{y_1}{x_2}} \frac{1}{\phi(\tilde{\Phi}(u_1))} \quad \dots.(**)$$

Since pdf Φ is standard normal distribution, then the equation (**) is probably positive. Then, we get the conditional distribution function formula from AR-N.GARCH model error for $F(\varepsilon_3|\varepsilon_2\ \varepsilon_1)$.

$$F(\varepsilon_3|\varepsilon_2\ \varepsilon_1) = \frac{\frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2, u_3)}{c(u_1, u_2)}$$

$$= \frac{\int_{-\infty}^{\Phi^{-1}(u_3)} f(\Phi^{-1}(u_1), \Phi^{-1}(u_2), s_3) ds_3 \frac{\partial \Phi^{-1}(u_1)}{\partial u_1} \cdot \frac{\partial \Phi^{-1}(u_2)}{\partial u_2}}{\exp\left(-\frac{1}{2}(y_1\ \ y_2)(\Lambda^{-1} - \mathrm{I})\binom{y_1}{y_2}\right)|\Lambda|^{-\frac{1}{2}}}$$

$$= \frac{\int_{-\infty}^{y_3} f(y_1, y_2, s_3) ds_3 \frac{1}{\phi(y_1)} \cdot \frac{1}{\phi(y_2)}}{\exp\left(-\frac{1}{2}(y_1\ \ y_2)(\Lambda^{-1} - \mathrm{I})\binom{y_1}{y_2}\right)|\Lambda|^{-\frac{1}{2}}} \qquad \text{is positive}$$

$$= \frac{\int_{-\infty}^{y_3} \frac{f(y_1, y_2, s_3)}{h(y_1, y_2)} ds_3 \frac{h(y_1, y_2)}{\phi(y_1)\phi(y_2)}}{\exp\left(-\frac{1}{2}(y_1\ \ y_2)(\Lambda^{-1} - \mathrm{I})\binom{y_1}{y_2}\right)|\Lambda|^{-\frac{1}{2}}}$$

$$= \frac{\int_{-\infty}^{y_3} \frac{f(y_1, y_2, s_3)}{h(y_1, y_2)} ds_3 C_\Lambda(u_1, u_2)}{\exp\left(-\frac{1}{2}(y_1\ \ y_2)(\Lambda^{-1} - \mathrm{I})\binom{y_1}{y_2}\right)|\Lambda|^{-\frac{1}{2}}}$$

$$= \int_{-\infty}^{y_3} \frac{f(y_1, y_2, s_3)}{h(y_1, y_2)} ds_3 \tag{13}$$

Relationship between $f_X(x)$ pdf normal standard distribution with pdf copula.

$$c_\Lambda(u_1, u_2) = \frac{h(x,y)}{f(x)g(y)} = \frac{h(F^{-1}(u_1), F^{-1}(u_2))}{f(F^{-1}(u_1))g(F^{-1}(u_2))} = \frac{h(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))} \tag{14}$$

In general, let be known the random variable $Y_k,\ k = 1, \dots, n$. We want to know the predicted value of Y_k if the previous information is known. The conditional distribution function of Y can be written :

$$P(Y_k \le y_k | Y_1, \dots Y_{k-1}) = C_n\left(\left\{\frac{g_k\left(\{u_i^{(j)}\}_{i=1}^k\right)}{c_{k-1}\left(\{u_i^{(j)}\}_{i=1}^k\right)}\right\}_{j=1}^n\right) \tag{15}$$

With the form of g_k as follows:

$$g_k\left(\{u_i^{(j)}\}_{i=1}^k\right) = \frac{\partial^{k-1}}{\partial u_{k-1} \dots \partial u_1} C_k\left(\{u_i^{(j)}\}_{i=1}^k\right)$$

$$= \left[\frac{\partial^{k-1}}{\partial z_{k-1} \dots \partial z_1} \Phi_{0;\Lambda}\left(\{z_1\}_{i=1}^k\right)\right] \prod_{i=1}^{k-1} \frac{\partial z_i}{\partial u_i}$$

$$= \int_{-\infty}^{z_k} \phi_{0,\Lambda}(z_1, \dots, z_{k-1}, \alpha) d\alpha \left[\prod_{i=1}^{k-1} \frac{\partial z_i}{\partial u_i}\right]^{-1}$$

$$= \left[\prod_{i=1}^{k-1} \frac{\partial z_i}{\partial u_i}\right]^{-1} \int_{-\infty}^{z_k} \frac{\phi_{0,\Lambda}(z_1, \dots, z_{k-1}, \alpha)}{\phi_{0,\mathrm{A}}(z_1, \dots, z_{k-1})} \phi_{0,\mathrm{A}}(z_1, \dots, z_{k-1}) d\alpha$$

$$= \frac{\phi_{0,\Lambda}(z_1, \dots, z_{k-1}, \alpha)}{\prod_{i=1}^{k-1} \phi_{0,1}(z_i)} \int_{-\infty}^{z_k} \phi_{\mu',\sigma'}(\alpha) d\alpha$$

$$= c_A(u_1, \dots, u_{k-1}) \Phi_{\mu',\sigma'}(z_k) \tag{16}$$

where:

$$u_i = F_{Y_i}(Y_i)$$
$$z_i = \Phi_{0;1}{}^{-1}(u_k{}^{(j)})$$
$$\mu' = \boldsymbol{k}_k^t\big(\boldsymbol{k}(x_k, x_k)\big)^{-1}\left(\begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix} - \mu_k\right)$$
$$\sigma' = -\,\boldsymbol{k}_k^t K^{-1}\boldsymbol{k}_k - \boldsymbol{k}(x_k, x_k)$$

So, the conditional distribution function of Y becomes

$$P(Y_k \le y_k | Y_1, \dots Y_{k-1}) = C_n\left(\left\{\frac{g_k\left(\{u_i{}^{(j)}\}_{i=1}^k\right)}{c_{k-1}\left(\{u_i{}^{(j)}\}_{i=1}^k\right)}\right\}_{j=1}^n\right)$$
$$= C_n\left(\left\{\frac{c_{k-1;A}\left(\{u_i{}^{(j)}\}_{i=1}^k\right)\Phi_{\mu',\sigma'}(z_k)}{c_{k-1}\left(\{u_i{}^{(j)}\}_{i=1}^k\right)}\right\}_{j=1}^n\right)$$
$$= C_n\left(\left\{\Phi_{\mu',\sigma'}(\Phi_{0;1}{}^{-1}(u_k{}^{(j)}))\right\}_{j=1}^n\right) \tag{17}$$

By changing the distribution function $\Phi_{\mu;\Sigma}$ into standard normal distribution and using Gaussian copula then the equation becomes

$$P(Y_k \le y_k | Y_1, \dots Y_{k-1}) = C_n\left(\left\{\Phi_{\mu',\sigma'}(\Phi_{0;1}{}^{-1}(u_k{}^{(j)}))\right\}_{j=1}^n\right)$$
$$= C_n\left(\left\{\Phi_{0;1}\left(\frac{\Phi_{0;1}{}^{-1}(u_k{}^{(j)}) - \mu'}{\sigma'}\right)\right\}_{j=1}^n\right)$$
$$= \Phi_\Lambda\left(\left\{\Phi^{-1}\left(\Phi_{0;1}\left(\frac{\Phi_{0;1}{}^{-1}(u_k{}^{(j)}) - \mu'}{\sigma'}\right)\right)\right\}_{j=1}^n\right)$$
$$= \Phi_\Lambda\left(\left\{\frac{\Phi_{0;1}{}^{-1}(u_k{}^{(j)}) - \mu'}{\sigma'}\right\}_{j=1}^n\right) \tag{18}$$

With probability density function of Y:

$$P\left(\{Y_k = y_k\}_{i=1}^k\right) = \prod_{s=1}^k c_n\left(\left\{\frac{g_s\left(\{u_i{}^{(j)}\}_{i=1}^s\right)}{c_{s-1}(\{u_i{}^{(j)}\}_{i=1}^{s-1})}\right\}_{j=1}^n\right)\prod_{j=1}^n\left[\frac{c_s\left(\{u_i{}^{(j)}\}_{i=1}^s\right)}{c_{s-1}(\{u_i{}^{(j)}\}_{i=1}^{s-1})}\, f_{Y_s{}^{(j)}}(Y_s{}^{(j)})\right]$$
$$= \prod_{s=1}^k c_n\left(\left\{\frac{c_{s-1;A}\left(\{u_i{}^{(j)}\}_{i=1}^s\right)\Phi_{\mu',\sigma'}(z_s)}{c_{s-1}(\{u_i{}^{(j)}\}_{i=1}^{s-1})}\right\}_{j=1}^n\right)\prod_{j=1}^n\left[\frac{c_s\left(\{u_i{}^{(j)}\}_{i=1}^s\right)}{c_{s-1}(\{u_i{}^{(j)}\}_{i=1}^{s-1})}\, f_{Y_s{}^{(j)}}(Y_s{}^{(j)})\right]$$
$$= \prod_{s=1}^k c_n\left\{\Phi_{\mu',\sigma'}(z_s)\right\}_{j=1}^n \prod_{j=1}^n\left[\frac{c_s\left(\{u_i{}^{(j)}\}_{i=1}^s\right)}{c_{s-1}\left(\{u_i{}^{(j)}\}_{i=1}^{s-1}\right)}\, f_{Y_s{}^{(j)}}(Y_s{}^{(j)})\right]$$
$$= \prod_{j=1}^n exp\left(-\frac{1}{2}Y^t(\sigma_s'^{-1} - I)Y\right)|\sigma_s'|^{-1/2}\prod_{j=1}^n\left[\frac{c_s\left(\{u_i{}^{(j)}\}_{i=1}^s\right)}{c_{s-1}\left(\{u_i{}^{(j)}\}_{i=1}^{s-1}\right)}\, f_{Y_s{}^{(j)}}(Y_s{}^{(j)})\right] \tag{19}$$

where:

$Y = \{Y^{(j)}\}$

$Y^{(j)} = \frac{\Phi_{0;1}{}^{-1}(u_k{}^{(j)})-\mu'}{\sigma'}$ with j=1,2,…,n.

Also, probability density function from Y (Y is Independent):

$$P\left(\{Y_k = y_k\}_{i=1}^{k}\right) = \prod_{i=1}^{n}\left[c_s\left(\{u_i{}^{(j)}\}_{i=1}^{k}\right)\prod_{s=1}^{k} f_{Y_s{}^{(j)}}(Y_s{}^{(j)})\right] \quad (20)$$

According to Probability Integral Transformation (PIT), if X a continuous random variable with the cumulative distribution function F_X then $Y = F_X(X) \sim Uniform(0,1)$. It means the transformation is indeed following the original distribution of the data, the set of values of the corresponding cumulative distribution function will be close to the distribution of Uniform (0.1. Also, to check the distribution function is following uniform (0.1). We are using Kolmogorov Smirnov test

$$KS_{data} = \frac{\sum_{i=1}^{T}(F_{est}(x_i)-F_{em}(x_i))^{\,2}}{T} \; ; \quad F_{em}(x) = \frac{\sum_{i=1}^{T}(I_{x_i \le x})}{T+1} \quad (21)$$

F_{est} is the distribution to be tested for compatibility

Then constructed a confidence interval in the following way:

for i:= 1 : $N_{interval}$

Generate data of $N_{interval}$ with size N_{data} node of the estimated distribution with the estimate parameter, suppose the data is

$w_{i,j}, j = 1, 2, \dots, N_{data}$

Calculate $KS(i) = \frac{\sum_{j=1}^{N_{data}}(F_{est}(w_{i,j})-F_{em}(w_{i,j}))^{\,2}}{N_{data}}$

endfor

Sort KS(1) to KS($N_{interval}$), select upper limit and lower limit based on interval levels according the significance of level. The magnitued of $N_{interval}$ is determined in such a way that the test results are stable / convergent.

4. The Step Construction of Copula Binding

4.1.1 Copula Univariate

1. Perform a fitting distribution to know the distribution of data errors.
   a. Estimated distribution parameters desired
   b. Distribution checking
      i. PIT (Probability Integral Transform)
      ii. Kolmogorov Smirnov
   c. If the distribution test results that the distribution does not match the data, then repeat process a.
2. Estimated kernel parameters

a. Select h_j ($h_j = j * 0.01, j = 1, 2, \dots, n$

b. Get the kernel

$$\Lambda_T(p,q) = h_j\, e^{\frac{|x_p - x_q|^2}{h_j}} \text{ dengan } p, q = 1,2, \dots, T$$

c. Determine the lower triangular matrix L by Cholesky factorization on the Λ_T matrix obtained above.

$$\Lambda_T = LL'$$

d. Complete the equation $Lv = z$.

With $z_j = \phi^{-1}(u_j)$, $u_j = F(x_j)$ in this case x_j is our data $\epsilon_{i,j}$

Calculate $L_j = -\sum_{j=1}^{T} \log L_{jj} + \frac{1}{2} z'z - \frac{1}{2} v'v$

e. Select h with criteria has the largest L_j

f. Using the estimated of h obtained in step 5 to get the kernel

3. Search for the cumulative distribution function of the marginal function by utilizing Copula Gaussian.

   Suppose $K_{k+1} = \begin{pmatrix} K_k & \boldsymbol{k}_k \\ \boldsymbol{k}_k^t & \boldsymbol{k}(x_k, x_k) \end{pmatrix}$ and $\mu_{k+1} = \begin{pmatrix} \boldsymbol{\mu}_k \\ \mu_{k+1} \end{pmatrix}$

   $\mu' = \boldsymbol{k}_k^t \left(\boldsymbol{k}(x_k, x_k)\right)^{-1} \left(\begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix} - \boldsymbol{\mu}_k \right)$

   $\sigma' = -\boldsymbol{k}_k^t K_k^{-1} \boldsymbol{k}_k - \boldsymbol{k}(x_k, x_k)$

   $\tilde{u}_k = F\left(\epsilon_{1,k} \middle| \Omega_{k-1}\right) = \Phi_{\mu',\sigma'}\left(\Phi_{0;1}{}^{-1}(u_k{}^{(j)})\right)$

   In the same way, $\tilde{v}_k = F\left(\epsilon_{2,k} \middle| \Omega_{k-1}\right)$

4. Testing the PIT (by utilizing the KS test for uniform distribution) on the marginal cumulative distribution function obtained to determine the congruent compatibility of Gaussian used. If not suitable, other Copulas may be used.

5. Looking at the characteristics of pair data $\tilde{u}_k$ and $\tilde{v}_k$. One way is to look at the scatter plot. After that can be predicted what kind of copula suitable to be binding copula.

4.1.2 Copula Bivariate

For example, Copula gauss is used. From the data of $\tilde{u}_k$ and $\tilde{v}_k$, we can obtain the parameters of Kendall's Tau (τ) and Pearson's Correlation Coefficient(ρ)

For $j = 1, 2, \ldots, n$ do the following:

1. Generate as many $n_{randomnumber}$ with pairs $\widehat{\tilde{u}_t} = F\left(\epsilon_{1,t} \middle| \Omega_{t-1}\right)$ and $\widehat{\tilde{v}_t} = G\left(\epsilon_{2,t} \middle| \Omega_{t-1}\right)$ which distributed Copula Gauss with parameter ρ.
2. Find expectation, upper also lower quartile to gain prediction expectations and 95% confidence interval
3. Perform the following process for the values obtained at no 2:
   a. Calculate $\widehat{z_{1,t}} = \Phi_{0,1}^{-1}(\widehat{\tilde{u}_t})$ and $\widehat{z_{2,t}} = \Phi_{0,1}^{-1}(\widehat{\tilde{v}_t})$
   b. Calculate $\widehat{\epsilon_{i,t}} = t_{\mu_{\epsilon_i}, \sigma_{\epsilon_i}, \nu_{\epsilon_i}}^{-1}\left(\Phi_{0,1}^{-1}\left(\Phi_{\mu',\sigma'}^{-1}\left(\widehat{z_{i,t}}\right)\right)\right)$ for $i = 1, 2$

      with $\mu' = \boldsymbol{k}_{\boldsymbol{i}_k}^t \left(\boldsymbol{k}_{\boldsymbol{i}}(x_k, x_k)\right)^{-1} \left(\begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix} - \mu_k \right)$

      and $\sigma' = -\boldsymbol{k}_{\boldsymbol{i}_k}^t K_i^{-1} \boldsymbol{k}_{\boldsymbol{i}_k} - \boldsymbol{k}_{\boldsymbol{i}}(x_k, x_k)$

      with K_1 matrix kernel for data 1 and K_2 for data 2
   c. Using AR(1)-N.GARCH(1,1) to obtain a predicted value $Y_t = \mu + \alpha Y_{t-1} + \sigma_t \epsilon_t$, $\epsilon_t \sim N(0,1)$

      $$\sigma_t^2 = \alpha_0 + \beta_1 \left(Y_{t-1} - \mu - \alpha Y_{(t-2)}\right)^2 + \beta_2 \sigma_{t-1}^2$$
   d. Obtained expectations and 95% of confidence interval

5. CONCLUSION

In this work, we have explained about compact model Copula Gaussian and AR(1)-N.GARCH(1,1). Also, we give illustrated to build the model based on a combination of traditional time series ARIMA with copula. As we know that time series with heavy-tail marginal distribution and complex underlying dependency structure is urgently needed, especially in the area of financial. Further, the estimation of model parameters must be relatively accurate and straightforward to allow this tool to be used in a high-throughput. At the same time, copula functions allows the specification and estimation of the dependency structure to be easily separated from those of the marginal distributions. In a nutshell, Our approach is flexible intuitive. It allows us to study covariate effects on the strength of dependence directly.

References

Caraka, R. ., Supari, & Tahmid, M. (2018). Copula-Based Model for Rainfall and El- Niño in Banyuwangi Indonesia. *Journal of Physics: Conference Series (JPCS)*, *1008*(1). https://doi.org/:10.1088/1742-6596/1008/1/012025

Caraka, R. ., Yasin, H., Sugiyarto, W., Sugiarto, & Ismail, K. M. (2016). Time Series Analysis Using Copula Gauss and AR(1)-N.GARCH(1,1). *MEDIA STATISTIKA*, *9*, 01–13. https://doi.org/DOI: 10.14710/medstat.9.1.1-13

Choroś, B., Ibragimov, R., & Permiakova, E. (2010). Copula Estimation. *Copula Theory and Its Applications SE - 3*, *198*, 77–91. https://doi.org/10.1007/978-3-642-12465-5_3

Genest, C., & Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, *12*(4), 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Kazianka, H. (2013). SpatialCopula: A Matlab toolbox for copula-based spatial analysis. *Stochastic Environmental Research and Risk Assessment*, *27*(1), 121–135. https://doi.org/10.1007/s00477-012-0571-3

Kolev, N., & Paiva, D. (2009). Copula-based regression models: A survey. *Journal of Statistical Planning and Inference*, *139*(11), 3847–3856. https://doi.org/10.1016/j.jspi.2009.05.023

Ma, J., & Sun, Z. (2011). Mutual information is copula entropy. *Tsinghua Science and Technology*, *16*(1), 51–54. https://doi.org/10.1016/S1007-0214(11)70008-6

Patton, A. (2013). Copula methods for forecasting multivariate time series. *Handbook of Economic Forecasting*, *2*, 899–960. https://doi.org/10.1016/B978-0-444-62731-5.00016-6