

TIME SERIES ANALYSIS USING COPULA GAUSS AND AR(1)-N.GARCH(1,1)

Rezzy Eko Caraka¹, Hasbi Yasin², Wawan Sugiarto¹, Sugiarto³, Kadi Mey Ismail¹

¹ Awardee of LPDP Scholarship, Ministry of Finance, Indonesia

² Department of Statistics, Diponegoro University, Indonesia

³ Senior High School 1 Moro, Riau Islands Province, Indonesia

e-mail: rezzyekocaraka@gmail.com

DOI: 10.14710/medstat.9.1.1-13

Abstract

In this case, the Gaussian Copula is used to connect the data that correlates with the time and with other data sets. Most often, practitioners rely only on the linear correlation to describe the degree of dependence between two or more variables; an approach that can lead to quite misleading conclusions as this measure is only capable of capturing linear relationships. Correlation doesn't mean causation, prediction using Copula is built on three things that the marginal distribution function, the kernel function, and the function of the Copula. Gaussian Copula involves the covariance matrix are approximated by using kernel functions. Kernel acts as the correlation between the approach of the data values that have the same characteristics. In this case, the characteristics used is the time. The advantage of the kernel function is able to calculate the correlation between random variables that have a realization using data characteristics. The advantage of using the kernel based Copula able to capture the dependencies between data and process data that have the same characteristics with time. Another benefit is that it allows a sequence of random variables have a joint distribution function so that the conditional probability of the prediction can be calculated.

Keywords: *Binding, Copula, GARCH, Gauss, Time Series*

1. PENDAHULUAN

Karakteristik umum harga saham adalah memiliki tingkat ketidakpastian. Ketidakpastian ini dalam hal pergerakan harga saham dalam jangka pendek, ataupun dalam jangka panjang. Karakteristik ini tidak disukai oleh para *investor* karena menimbulkan risiko pada investasi mereka. Ketidakpastian ini juga tidak dapat dihindari dalam investasi. Untuk menghadapi ketidakpastian pergerakan harga, yang dapat dilakukan oleh *investor* adalah mengurangi ketidakpastian tersebut. Salah satu alat (*tools*) untuk mengurangi ketidakpastian adalah dengan melakukan prediksi (*forecasting*) harga saham.

Proses prediksi harga saham terbagi menjadi dua kelompok. Kelompok pertama adalah pihak yang mempercayai bahwa terdapat suatu cara untuk memprediksi harga saham. Kelompok kedua adalah pihak yang berkeyakinan bahwa pasar adalah efisien, dan apabila ada informasi baru yang diperoleh, pasar akan menyerapnya dan mengoreksi dirinya sendiri. Kelompok kedua ini percaya bahwa harga saham tidak dapat diprediksi. Hal inilah yang disebut teori *Efficient Market Hypothesis* (EMH) seperti yang disebutkan oleh Gryc (2006). Kelompok ini beranggapan bahwa pasar saham mengikuti pola *random*

walk, yang berarti bahwa prediksi terbaik yang dapat diperoleh tentang harga masa depan suatu saham adalah berdasarkan harga saham saat ini.

Penelitian mengenai fluktuasi saham yang dilakukan oleh Caraka dan Yasin (2015) menyimpulkan *capital gain* dapat dipakai sebagai salah satu acuan dalam melakukan investasi saham dengan melihat besarnya kerugian maupun keuntungan yang mungkin akan dialami oleh seorang investor terutama jika ingin melakukan investasi dalam jangka waktu yang pendek. Harga saham mengalami fluktuasi baik berupa kenaikan maupun penurunan sehingga dengan harga saham yang berfluktuasi memberikan peluang kepada para investor untuk mengalami keuntungan maupun kerugian. Salah satu metode dalam memprediksi harga saham adalah metode *time series forecasting*. Pada metode ini dicoba dibuat model prediksi linear untuk melihat pola dari data historis harga saham untuk menilai harganya di masa depan. Nilai harga saham di masa depan dianggap sebagai kombinasi linear dari data historisnya. Model linear ini terbagi menjadi dua kategori model, yakni model *univariate regression* dan model *multivariate regression*. Beberapa yang digunakan pada metode *time series forecasting* diantaranya AR (p), GARCH (p,q), dan *Copula*. Ada banyak model *Copula* yang bisa digunakan diantaranya *Copula Gaussian*, *Copula-t*, *Copula Clayton*, *Copula Frank*, dan *Copula Gumbel*.

Copula merupakan salah satu metode statistika yang dapat menggambarkan hubungan antar variabel yang tidak terlalu ketat terhadap asumsi distribusi, serta dapat menunjukkan hubungan dependensi pada titik-titik ekstrim dengan jelas. Metode ini mempunyai kemampuan untuk mendeskripsikan struktur dependensi antar variabel dengan marginal yang berbeda dan memodelkan dependensi *tail*-nya. *Copula* adalah suatu fungsi dari dua hubungan distribusi yang masing-masing mempunyai fungsi marginal distribusi (Nelsen, 1998). Beberapa penelitian mengenai *Copula* telah dilakukan, antara lain penelitian oleh Murteira dan Lourenço (2007) mengenai penggunaan *Copula* pada kasus kesehatan. Zhu, Ghosh, dan Goodwin (2008) menerapkan *Copula* untuk memodelkan asuransi.

2. TINJAUAN PUSTAKA

2.1 Konsep Korelasi

Analisis korelasi adalah metode statistik yang digunakan untuk mengukur besarnya hubungan linier antara dua variabel dengan satu nilai yang dinamakan koefisien korelasi (Walpole, 2007). Nilai korelasi berkisar pada interval $-1 \leq \rho \leq 1$. Jika korelasi positif, maka hubungan antara dua variabel adalah searah. Jika korelasi bernilai negatif, maka hubungan antara dua variabel adalah berlawanan. Taksiran ρ yang dekat ke satu, berarti adanya korelasi yang kuat/tinggi atau ikatan linier, sedangkan nilai dekat dengan nol menunjukkan korelasi yang lemah atau tidak ada korelasi. Berdasarkan skala pengamatan pada data pengamatan, korelasi dibedakan sebagai berikut:

1. Korelasi Pearson untuk mengetahui hubungan antara dua variabel yang mempunyai skala interval atau rasio (untuk statistik parametrik).
2. Korelasi Tau-Kendall dan korelasi Rank Spearman untuk mengetahui hubungan antara dua variabel yang memiliki skala ordinal (untuk statistik nonparametrik).

2.2 Copula

Copula adalah suatu fungsi distribusi bersama dari beberapa fungsi distribusi marjinal. *Copula* digunakan untuk menganalisis ketergantungan variable-variabel acak dalam struktur yang digambarkan oleh fungsi gabungan tersebut (Nelsen, 1998).

2.2.1 Copula Gauss

Berikut adalah *Copula Gauss* untuk dua variabel, yang dapat dengan mudah diperluas ke multivariabel (Schölzel & Friederichs, 2008).

Misalkan $(X, Y) \sim$ Bivariat Normal dan $X \sim N(0,1), Y \sim N(0,1)$

$$h(x, y) = f(x) \cdot g(y) \cdot c(u, v), \text{ dengan } u = F(x), v = G(y)$$

Pdf Gaussian *Copula* dinyatakan dalam persamaan (1)

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

dengan $x = \Phi^{-1}(u)$, $y = \Phi^{-1}(v)$, dan $\Phi(x)$ adalah distribusi kumulatif dan didefinisikan sebagai normal standar. Maka Gaussian *Copula* dapat didefinisikan sebagai:

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

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

$$\Phi_{0,1}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

Misal $u_i = F_i(x_i)$, sehingga dapat dinyatakan sebagai:

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

Jika $F_i(\cdot) \sim \Phi_{\mu_i, \sigma_i^2}(\cdot)$, maka :

$$C_{\Lambda}(u_1, \dots, u_T) = \Phi_{\Lambda}(\Phi_{0,1}^{-1}(\Phi_{\mu_1, \sigma_1^2}(x_1)), \Phi_{0,1}^{-1}(\Phi_{\mu_2, \sigma_2^2}(x_2)), \dots, \Phi_{0,1}^{-1}(\Phi_{\mu_T, \sigma_T^2}(x_T)))$$

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

misalkan $X_i \sim F$

Sehingga *Copula Gauss* dapat dinyatakan sebagai berikut.

$$C_{\Lambda} = \Phi_{\Lambda}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_T)), \quad (2)$$

dengan Φ_{Λ} adalah fungsi distribusi kumulatif bersama normal multivariat dengan matriks kovariansi Λ , Φ^{-1} invers cdf univariat normal baku, dan $u_i = F(x_i)$. Selain itu *Probability density function* (pdf) dari $F(x)$ adalah $f_x(x)$ dinyatakan dalam persamaan (3) yaitu:

$$f_x(x) = \frac{1}{(2\pi)^{\frac{T}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (3)$$

dengan $x' = (x_1, x_2, \dots, x_T)$

$$\Sigma = \begin{bmatrix} \text{var}(x_1, x_1) & \text{cov}(x_1, x_2) & \cdots & \text{cov}(x_1, x_T) \\ \vdots & \ddots & \vdots & \vdots \\ \text{cov}(x_T, x_1) & \cdots & \cdots & \text{var}(x_T, x_T) \end{bmatrix}$$

Selanjutnya didapat:

$$F_X(y_1, y_2, \dots, y_T) = \int_{-\infty}^{y_T} \int_{-\infty}^{t_{T-1}} \cdots \int_{-\infty}^{y_1} f(x) dx_1 dx_2 \cdots dx_T$$

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

2.2.2 Copula Gauss untuk Binding

Copula binding dilakukan untuk menggabungkan dua atau lebih *Copula*. Khusus untuk *Copula Gauss* Bivariat, pdf dari *Copula* bisa ditulis dalam persamaan (4)

$$\begin{aligned} c(u, v; \rho) &:= \frac{\partial^2}{\partial u \partial v} C(u, v; \rho) = \frac{\varphi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho)}{\varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v))} \\ &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\rho\Phi^{-1}(u)\Phi^{-1}(v) - \rho^2\Phi^{-1}(u)^2\Phi^{-1}(v)^2}{2(1-\rho^2)}\right) \end{aligned} \quad (4)$$

dengan cdf (*Cumulative distribution function*) dapat dinyatakan sebagai berikut

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

2.2.3 Prediksi Menggunakan Copula Gauss

Dalam melakukan prediksi menggunakan *copula gauss* dimiliki formula fungsi distribusi bersyarat dari error model AR-N.GARCH untuk $F(\varepsilon_2 | \varepsilon_1)$.

$$\begin{aligned} F(\varepsilon_2 | \varepsilon_1) &= \frac{\frac{\partial}{\partial u_1} C(u_1, u_2)}{c(u_1) = 1} = \int_{-\infty}^{\Phi^{-1}(u_2)} f(\Phi^{-1}(u_1), s_2) ds_2 \frac{\partial \Phi^{-1}(u_1)}{\partial u_1} \\ &= \int_{-\infty}^{\Phi^{-1}(u_2)} f(\Phi^{-1}(u_1), s_2) ds_2 \frac{1}{\phi(\tilde{\Phi}(u_1))} \end{aligned} \quad (5)$$

Misalkan $y_1 = \Phi^{-1}(u_1)$, maka persamaan (5) dapat didefinisikan menjadi:

$$\int_{-\infty}^{y_1} \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_1 \ x_2) \Sigma^{-1} \begin{pmatrix} y_1 \\ x_2 \end{pmatrix}} \frac{1}{\phi(\tilde{\Phi}(u_1))} \quad (6)$$

Karena pdf Φ berdistribusi normal baku, maka persamaan (5) dipastikan bernilai positif. Kemudian, diperoleh formula fungsi distribusi bersyarat dari error model AR-N.GARCH untuk $F(\varepsilon_3 | \varepsilon_2 \ \varepsilon_1)$.

$$\begin{aligned}
F(\varepsilon_3 | \varepsilon_2 \varepsilon_1) &= \frac{\frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2, u_3)}{c(u_1, u_2)} \\
&= \frac{\int_{-\infty}^{\Phi^{-1}(u_3)} f(\Phi^{-1}(u_1), \Phi^{-1}(u_2), s_3) ds_3 \frac{\partial \Phi^{-1}(u_1)}{\partial u_1} \cdot \frac{\partial \Phi^{-1}(u_2)}{\partial u_2}}{\exp\left(-\frac{1}{2}(y_1 \ y_2)(\Lambda^{-1} - I)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) |\Lambda|^{-\frac{1}{2}}} \\
&= \frac{\int_{-\infty}^{y_3} f(y_1, y_2, s_3) ds_3 \frac{1}{\phi(y_1)} \cdot \frac{1}{\phi(y_2)}}{\exp\left(-\frac{1}{2}(y_1 \ y_2)(\Lambda^{-1} - I)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) |\Lambda|^{-\frac{1}{2}}} \text{ bernilai positif.} \\
&= \frac{\int_{-\infty}^{y_3} \frac{f(y_1, y_2, s_3)}{h(y_1, y_2)} ds_3 \frac{h(y_1, y_2)}{\phi(y_1)\phi(y_2)}}{\exp\left(-\frac{1}{2}(y_1 \ y_2)(\Lambda^{-1} - I)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) |\Lambda|^{-\frac{1}{2}}} \\
&= \frac{\int_{-\infty}^{y_3} \frac{f(y_1, y_2, s_3)}{h(y_1, y_2)} ds_3 C_\Lambda(u_1, u_2)}{\exp\left(-\frac{1}{2}(y_1 \ y_2)(\Lambda^{-1} - I)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) |\Lambda|^{-\frac{1}{2}}} \\
&= \int_{-\infty}^{y_3} \frac{f(y_1, y_2, s_3)}{h(y_1, y_2)} ds_3
\end{aligned}$$

Hubungan antara $f_X(x)$ pdf distribusi normal baku dengan pdf *Copula* dapat dinyatakan sebagai berikut:

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

Misalkan diketahui peubah acak Y_k , $k = 1, \dots, n$ ingin mengetahui prediksi nilai Y_k maka fungsi distribusi bersama bersyarat dari Y yaitu:

$$P(Y_k \leq y_k | Y_1, \dots, Y_{k-1}) = C_n \left(\left\{ \frac{g_k(\{u_i^{(j)}\}_{i=1}^k)}{c_{k-1}(\{u_i^{(j)}\}_{i=1}^k)} \right\}_{j=1}^n \right) \quad (7)$$

Dengan bentuk g_k pada persamaan (7) seperti berikut :

$$\begin{aligned}
g_k(\{u_i^{(j)}\}_{i=1}^k) &= \frac{\partial^{k-1}}{\partial u_{k-1} \dots \partial u_1} C_k(\{u_i^{(j)}\}_{i=1}^k) \\
&= \left[\frac{\partial^{k-1}}{\partial z_{k-1} \dots \partial z_1} \Phi_{0, \Lambda}(\{z_i\}_{i=1}^k) \right] \prod_{i=1}^{k-1} \frac{\partial z_i}{\partial u_i}
\end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{z_k} \Phi_{0,\Lambda}(z_1, \dots, z_{k-1}, \alpha) d\alpha \left[\prod_{i=1}^{k-1} \frac{\partial z_i}{\partial u_i} \right]^{-1} \\
&= \left[\prod_{i=1}^{k-1} \frac{\partial z_i}{\partial u_i} \right]^{-1} \int_{-\infty}^{z_k} \frac{\Phi_{0,\Lambda}(z_1, \dots, z_{k-1}, \alpha)}{\Phi_{0,A}(z_1, \dots, z_{k-1})} \Phi_{0,A}(z_1, \dots, z_{k-1}) d\alpha \\
&= \frac{\Phi_{0,\Lambda}(z_1, \dots, z_{k-1}, \alpha)}{\prod_{i=1}^{k-1} \Phi_{0,1}(z_i)} \int_{-\infty}^{z_k} \Phi_{\mu',\sigma'}(\alpha) d\alpha \\
&= c_A(u_1, \dots, u_{k-1}) \Phi_{\mu',\sigma'}(z_k)
\end{aligned}$$

dengan

$$u_i = F_{Y_i}(Y_i)$$

$$z_i = \Phi_{0;1}^{-1}(u_k^{(j)})$$

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

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

Sehingga fungsi distribusi bersyarat dari Y menjadi:

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

Dengan mengubah fungsi distribusi $\Phi_{\mu;\Sigma}$ menjadi distribusi normal baku dan menggunakan *Copula Gaussian* maka persamaan menjadi:

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

Fungsi kepadatan peluang bersama dari Y dapat dinyatakan:

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

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

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

Fungsi kepadatan peluang bersama dari Y (jika Y saling bebas):

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

2.3 Uji Kecocokan Data menggunakan PIT dan Uji KS

Menurut *Probability Integral Transformation* (PIT), jika X suatu peubah acak kontinu dengan fungsi distribusi kumulatif F_X maka $Y = F_X(X) \sim Uniform(0,1)$. Artinya adalah jika transformasi memang sesuai dengan distribusi asli dari data, kumpulan nilai-nilai fungsi distribusi kumulatif yang bersangkutan akan mendekati distribusi Uniform(0,1). Untuk menguji apakah histogram dari nilai-nilai fungsi distribusi kumulatif berdistribusi Uniform(0,1) atau tidak, dapat digunakan uji KS.

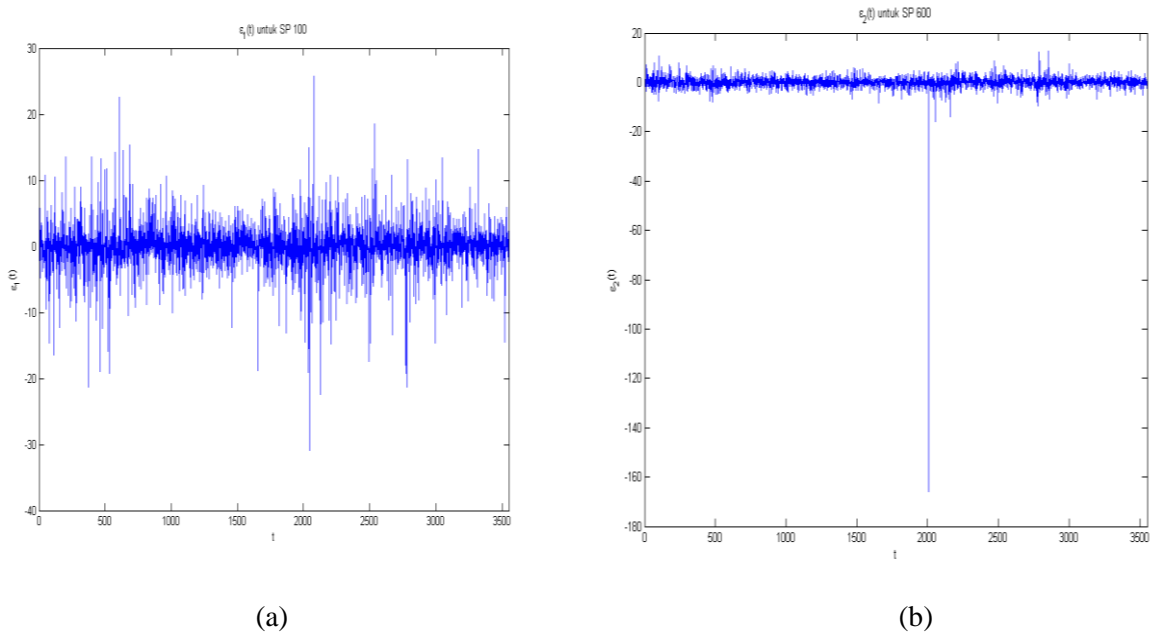
3. METODOLOGI PENELITIAN

S&P 100 adalah pasar saham Amerika Serikat yang dikelola oleh Standard & Poor. S&P 100 bagian dari S&P 500 yang termasuk 100 saham AS terkemuka. Konstituen dari S & P 100 yang dipilih untuk keseimbangan sektor dan mewakili sekitar 57% dari kapitalisasi pasar S&P 500 dan hampir 45% dari kapitalisasi pasar dari pasar ekuitas AS. Sedangkan S&P 600 adalah istilah lain untuk indeks S&P *small-cap* 600 dan merupakan indeks saham yang berpengaruh pada perekonomian Amerika. Indeks saham ini diisi oleh saham-saham dari sekitar 600 perusahaan yang memiliki modal besar di Amerika Serikat. Indeks saham ini dikelola oleh dan di *manage* oleh Standard and Poor's, sebuah divisi keuangan dari McGraw-Hill yang merupakan perusahaan peringkat atas saham dan obligasi. Dalam artikel ini, harga saham S&P 100 dan S&P 600 dimodelkan menjadi data *time series*, yang akan juga melibatkan *Copula* Gauss Multivariat.

4. HASIL DAN PEMBAHASAN

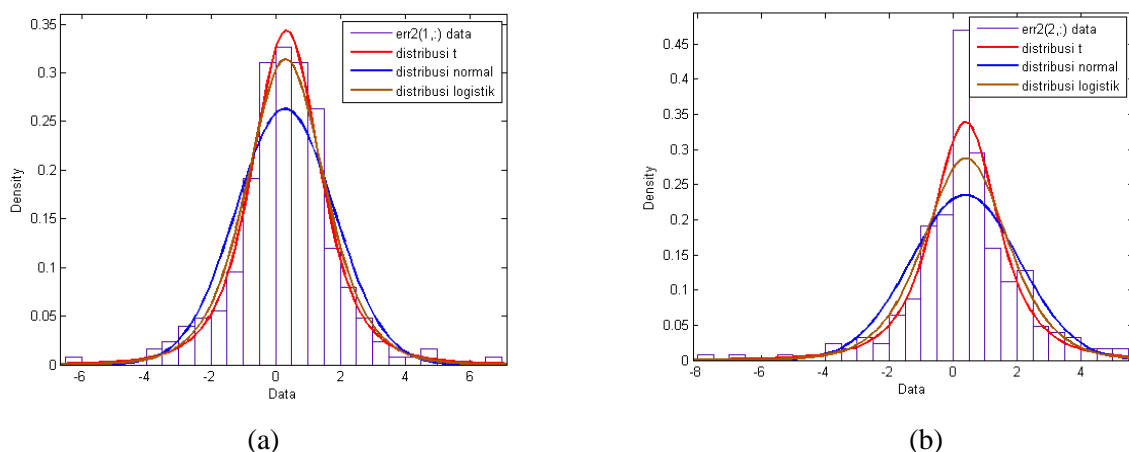
4.1 Fungsi Distribusi Marginal dari Error Model AR(1)-N.GARCH(1,1)

Berikut adalah plot error $\epsilon_{1,t}$ dan $\epsilon_{2,t}$ yang diperoleh dari model AR(1)-N.GARCH(1,1)



Gambar 1. Plot Data (a). $\epsilon_{1,t}$ AR(1)-N.GARCH(1,1), (b). $\epsilon_{2,t}$ AR(1)-N.GARCH(1,1)

Selanjutnya dilakukan *fitting* distribusi pada $\epsilon_{i,t}$ untuk mengetahui fungsi marginalnya yang diperoleh dari model AR-N.GARCH.



Gambar 2. Plot Data (a).Fitting Distribusi Marginal Error SP100, (b). Fitting Distribusi Marginal Error SP600

Dapat dilihat bahwa secara grafis distribusi t diduga cukup sesuai dengan $\epsilon_{1,t}$ dengan parameter sebagai berikut:

Tabel 1. Estimasi Parameter

Saham		$\hat{\mu}$	$\hat{\sigma}$	ν
Sp100	$\hat{\epsilon}_{1,t}$	0,327187055631610	1,088556910692113	3,801952514914956
Sp600	$\hat{\epsilon}_{2,t}$	0,385579999524267	1,079919890458092	2,871901167775360

Selanjutnya akan diuji hipotesis bahwa $\epsilon_{i,t}$ berdistribusi t dengan menggunakan metode KS untuk data berukuran besar. Statistik uji yang digunakan:

$$KS_{data} = \frac{\sum_{i=1}^T (F_{\mu,\sigma,\nu}(x_i) - F_{em}(x_i))^2}{T} \quad \text{dengan} \quad F_{em}(x) = \frac{\sum_{i=1}^T (I_{x_i \leq x})}{T+1}$$

Dalam hal ini, digunakan 1000 nilai KS untuk membangun selang kepercayaan, yang masing-masingnya dihitung dengan membangkitkan 1000 data dari distribusi taksiran yang telah diperoleh sebelumnya. Dengan menggunakan uji 1 arah, diperoleh selang kepercayaan dan statistik uji.

Tabel 2. Selang Kepercayaan dan Statistik Uji

	Data $\epsilon_{1,t}$	Data $\epsilon_{2,t}$
Selang kepercayaan	(1,001 x 10 ⁻⁴ ; 1,1935 x 10 ⁻⁴)	(1,360 x 10 ⁻⁴ ; 1,433 x 10 ⁻⁴)
Statistik uji	1,166 x 10 ⁻⁴	4,5032 x 10 ⁻⁴

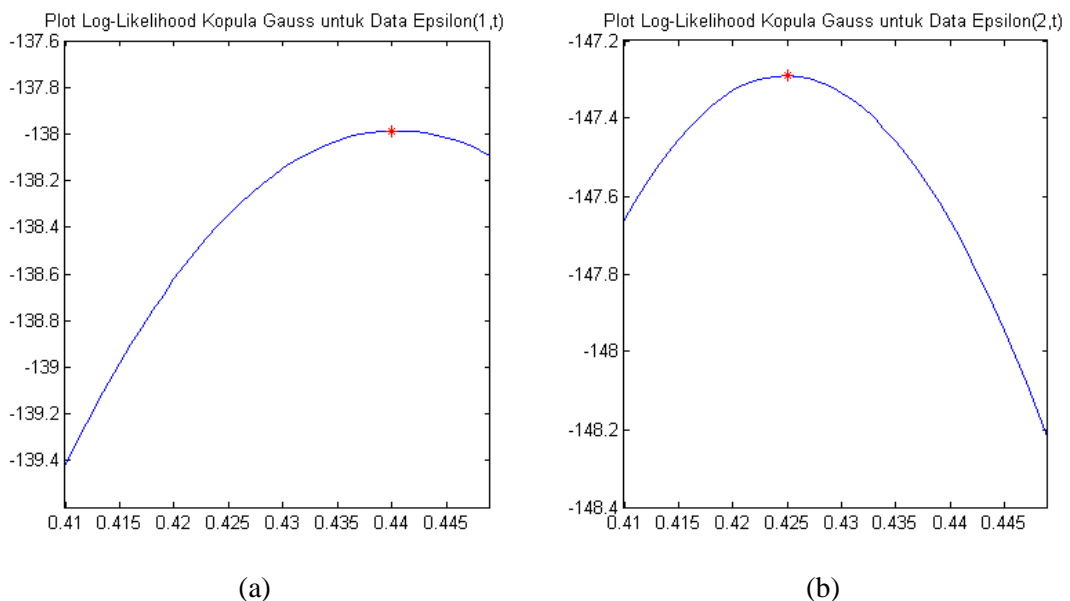
Nilai ini berada pada daerah selang kepercayaan. Karena itu, hipotesis bahwa distribusi dari $\epsilon_{i,t}$ berasal dari distribusi t dengan parameter yang bersangkutan tidak ditolak.

4.2 Penentuan Parameter *Copula* Gaussian

Diperoleh h yang memaksimumkan fungsi *likelihood* sebagai berikut. Dalam hal ini, digunakan partisi sebesar 0,001 untuk setiap iterasi pencarian nilai h

Tabel 3. Parameter *Copula* Gauss

Saham	h
S&P 100	0,440
S&P 600	0,425



Gambar 3. Plot Data (a). $\epsilon_{1,t}$ AR(1)-N.GARCH(1,1), (b). $\epsilon_{2,t}$ AR(1)-N.GARCH(1,1)

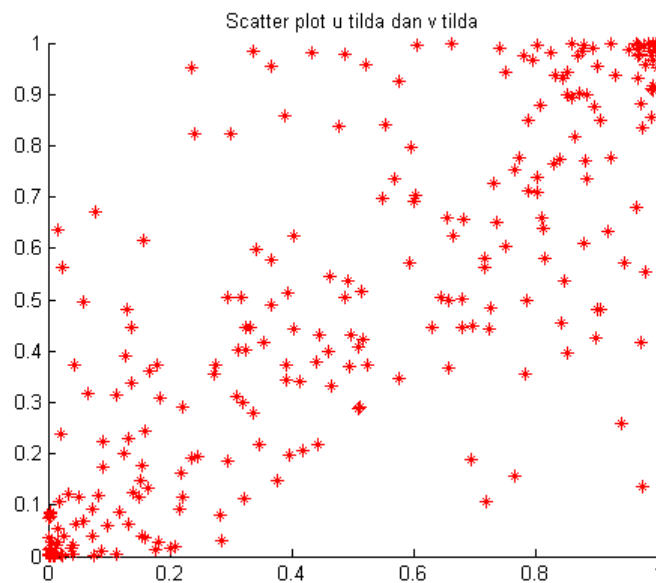
Misalkan $\epsilon_1, \dots, \epsilon_T \sim F$ adalah nilai error untuk S&P 100 dan $\delta_1, \dots, \delta_T \sim G$ adalah nilai error untuk S&P 600 yang diperoleh dari model GARCH. Dari *fitting* distribusi yang telah dilakukan menggunakan metode grafis maupun metode K-S diperoleh masing-masing F dan G adalah distribusi student's t . Selanjutnya akan ditentukan distribusi bersyarat nilai error dari model sebagai berikut.

$$F(\epsilon_T | \Omega_{T-1}) = \frac{\partial^{T-1}}{\partial u_1, \dots, u_{T-1}} \frac{C_\Lambda(u_1, \dots, u_T)}{C_\Lambda(u_1, \dots, u_{T-1})}$$

untuk mendapatkan Λ , digunakan h taksiran yang memaksimumkan likelihood yang telah diperoleh sebelumnya. Dengan formula yang sama juga didapatkan $G(\epsilon_T | \Omega_{T-1})$. Misalkan

$$\begin{aligned} \tilde{u}_1 = F(\epsilon_1 | \Omega_0) &= F(\epsilon_1) & \text{dan} & & \tilde{v}_1 = F(\delta_1 | \Omega_0) &= F(\delta_1) \\ \vdots & & & & \vdots & \\ \tilde{u}_T = F(\epsilon_T | \Omega_{T-1}) &= F(\epsilon_T) & & & \tilde{v}_T = F(\delta_T | \Omega_{T-1}) &= F(\delta_T) \end{aligned}$$

\tilde{u}_T dan \tilde{v}_T yang berupa distribusi normal dengan mean dan variansi yang diperbarui untuk setiap hari ke- k , $k = 1, 2, \dots, T$. Berikut adalah *scatter plot* untuk \tilde{u}_t dan \tilde{v}_t

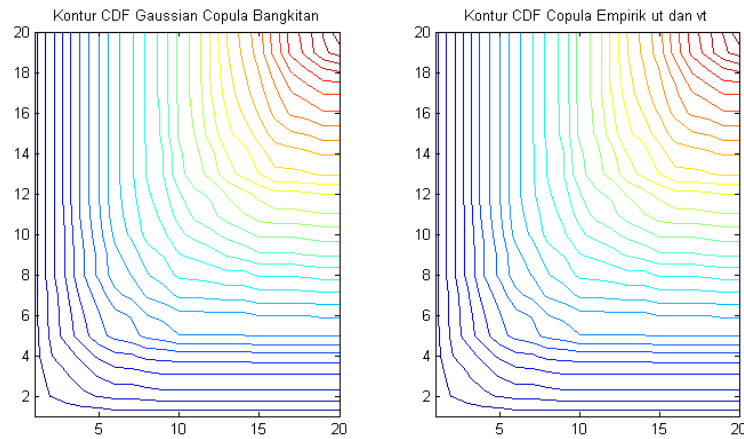


Gambar 4. Scatter Plot \tilde{u}_t dan \tilde{v}_t

Menurut *Probability Integral Transformation* (PIT), jika transformasi memang sesuai dengan distribusi asli dari data, kumpulan nilai-nilai fungsi distribusi kumulatif yang bersangkutan akan mendekati distribusi Uniform(0,1).

4.3 Copula Gauss untuk Binding

Dengan menggunakan data \tilde{u}_t dan \tilde{v}_t , diperoleh Koefisien Korelasi Kendall's Tau $\tau = 0,621418326693227$. Ini berarti bahwa data-data \tilde{u}_t dan \tilde{v}_t memiliki korelasi yang cukup kuat. Diperoleh pula koefisien korelasi Pearson $\rho = 0,828330787496002$ ini berarti nilai korelasi yang sangat kuat.

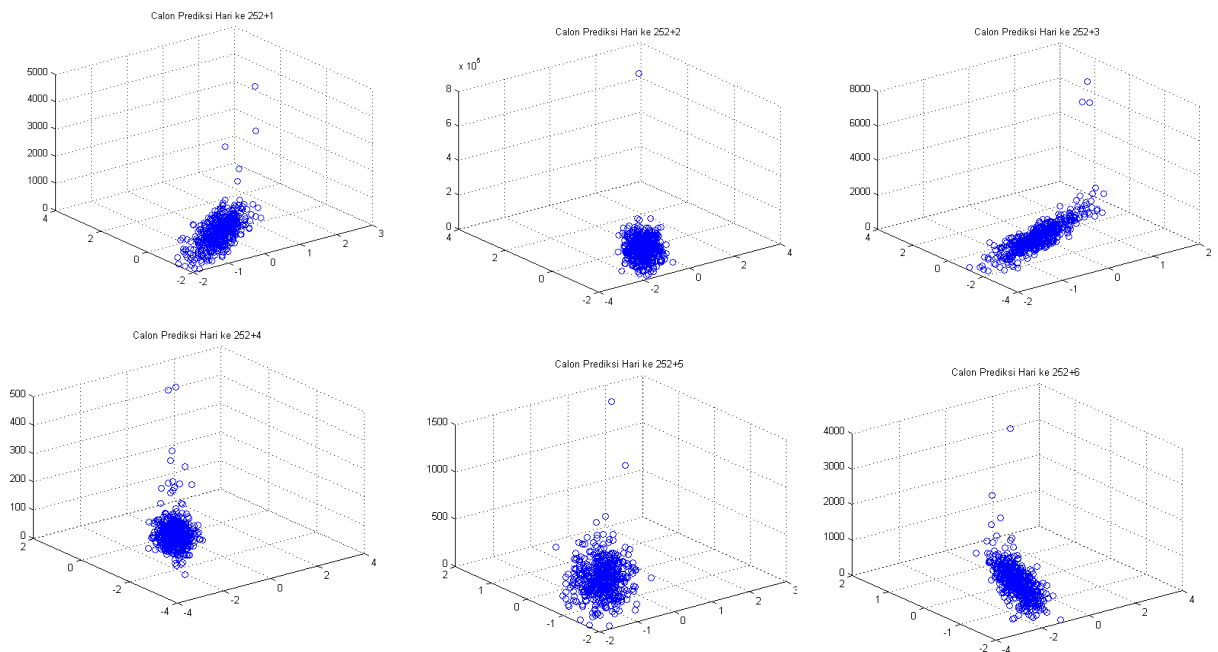


Gambar 5. Kontur CDF Gaussian *Copula*

Copula Gauss cukup dapat mewakili nilai-nilai data. Akan tetapi, bagian ekor di nilai yang mendekati (0,0) dan (1,1) tidak dapat terwakili oleh *Copula* Gauss. Lalu, dihitung pula nilai $C(\tilde{u}_t, \tilde{v}_t)$. Dengan uji yang sama dengan sebelumnya, akan diuji bahwa nilai-nilai dari $C(\tilde{u}_t, \tilde{v}_t)$ berdistribusi $uniform(0,1)$. Diperoleh selang kepercayaan 95%: (0,012896332819681; 0,014263766812160) dengan nilai statistik uji 0,013942682837266. Nilai statistik uji berada di dalam selang kepercayaan. Jadi hipotesis sebelumnya tidak ditolak. Artinya adalah bahwa berdasarkan PIT, data dari \tilde{u}_t dan \tilde{v}_t sesuai dengan *Copula* ini.

4.4 Prediksi Harga Saham

Dengan membangkitkan sejumlah $n_{bangkit} = 500$ nilai \tilde{u}_t dan \tilde{v}_t yang berdistribusi *Copula* Gauss dengan ρ seperti yang telah diperoleh, lalu menginverskannya, kemudian menghitung nilai peluang bersama menggunakan *Copula* Gauss *binding* dan multivariat, diperoleh *scatter* berikut:



Gambar 5. Scatter Plot Prediksi

Melihat adanya pencilan, untuk menentukan prediksi, tidak diambil pasangan nilai \hat{u}_t dan \hat{v}_t yang memaksimumkan nilai $h(u_t, v_t | \Omega_{t-1})$. Dengan batas atas dan batas bawah harga saham sebagai berikut:

Tabel 4. Batas Bawah dan Batas Atas serta Realisasi Harga Saham

Hari ke	Harga saham SP100			Harga saham SP600		
	Batas bawah prediksi	Batas atas prediksi	Realisasi	Batas bawah prediksi	Batas atas prediksi	Realisasi
1	818,2400	818,2400	818,2400	118,9000	118,9000	18,9000
2	815,3143	823,1736	819,1815	118,3883	119,8828	19,1056
3	814,1520	826,1092	820,1842	118,2700	120,3095	19,2714
4	816,9508	825,0372	821,0508	118,6826	120,2398	19,4556
5	817,8686	826,3063	821,9435	118,8427	120,5136	19,6509
6	817,9478	827,4406	822,7367	118,6375	121,0621	19,8699
7	819,6914	828,3053	823,6717	119,2001	120,9984	20,0663

5. KESIMPULAN

Prediksi menggunakan *Copula* ini dibangun atas tiga hal penting yaitu fungsi distribusi marjinal, fungsi kernel, dan fungsi *Copula*. Dalam hal ini, digunakan *Copula* Gaussian untuk menghubungkan data yang berkorelasi dengan waktu dan dengan himpunan data lainnya (dalam hal ini data return harga saham SP100 dan SP600).

Acknowledgment

Penulis mengucapkan terima kasih kepada Dr.Utriweni Mukhaiyar, M.Si, Ayu Shabrina, Sumayyah Roihanah Thoyyibah berkat masukan dan kontribusi. Penelitian ini didukung penuh oleh Lembaga Pengelola Dana Pendidikan (LPDP) Kementerian Keuangan Republik Indonesia.

DAFTAR PUSTAKA

- Annisa, K.N. dan Sutikno. 2015. Analisis Hubungan Curah Hujan dan Indikator El-Nino Southern Oscillation di Sentra Produksi Padi Jawa Timur dengan Pendekatan Copula. *Jurnal Sains dan Seni ITS*, Vol. 4, No.1, pp 2337-3520.
- Arshad, M., Rasool, M. T. dan Ahmad, M. 2003. Anderson Darling and Modified Anderson Darling Tests for Generalized Pareto Distribution. *Pakistan Journal of Applied Sciences*, 3(2), pp 85-88.
- Berg, D. dan Bakken, H. 2006. *Copula Goodness-of-fit Tests: A comparative Study*. The Norwegian Computing.
- Caraka, R.E., Yasin, H. dan Prahutama, A. 2015. Pemodelan General Regression Neural Network (GRNN) Pada Data Return Indeks Harga Saham Euro 50. *Journal Gaussian*, Universitas Diponegoro, Vol. 4, No. 2, pp 181-192.
- Caraka, R.E. dan Yasin, H. 2014. Pemodelan General Regression Neural Network (GRNN) Dengan Peubah Input data Return Untuk Peramalan Indeks Hang Seng. *Prosiding Seminar Nasional Ilmu Komputer (SNIK)*, Universitas Negeri Semarang, ISBN: 978-602-71550-0-8, pp 283-288
- Embrechts, P., Lindskog, F. dan McNeil, A. 2001. *Modelling Dependence with Copulas and Application to Risk Management*. Switzerland. Departement of Mathematics, ETHZ CH-8092 Zürich.

- Genest, C. dan Nešlehová, J. 2010. *Copulas: Introduction to the Theory and Implementation in R with Applications in Finance and Insurance*. Universite Laval and McGill University.
- Genest, C. dan Riverst, L. P. 1993. Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistics Association*, pp 1034-1043.
- Mikosch, T. 2008. *Copulas: Tales and Facts*. Denmark: Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100.
- Mukhaiyar, U. 2012. *Pengenalan Analisis Deret Waktu (Time Series Analysis)*. Catatan kuliah. ITB
- Murteira, J. M. dan Lourenço, Ó. D. 2007. *Health Care Utilization and Self-Assessed Health: Specification of Bivariate Models Using Copulas*. Health, Econometrics and Data Group.
- Nelsen, R. B. 1998. *An Introduction to Copulas Second Edition*. USA. Springer.
- Schölzel, C., dan Friederichs, P. 2008. Multivariate Non-Normally Distributed Random Variables In Climate Research – Introduction to The Copula Approach. *Nonlin. Processes Geophys*, 15, pp 761–772.
- Shabrina, A. 2014. *Analisis Model Proses Kopula Berdasarkan Kernel dan Aplikasinya*. Bandung. ITB.
- Walpole, R. E. 2007. *Probability dan Statistics for Engineers dan Scientists Eighth Edition*. London. Pearson Education LTD.
- Zhu, Y., Ghosh, S. K. dan Goodwin, B. K. 2008. *Modeling Dependence in the Design of Whole Insurance Contract -A Copula-Based Model Approach*.