Gaussian Copula Regression Application

Samia A. Adham
Department of Statistics, Faculty of Science
King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia

Mervat K. Abd Elaal
Department of Statistics, Faculty of Science
King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia
&
Department of Statistics, Faculty of Commerce
Al-Azhar University, Girls Branch
Ciro, Egypt

Heba M. Malaka
Department of Statistics, Faculty of Science
King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia

Copyright © 2016 Samia A. Adham et al. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract
Gaussian copula models are frequently used to extend univariate regression models to the multivariate case. The main benefit of the topic is that the specification of the regression model is conveniently separated from the dependence structure described in the familiar form of the correlation matrix of a multivariate Gaussian distribution [1]. This form of flexibility has been successfully employed in several complex applications including longitudinal data analysis, spatial statistics, genetics and time series.
In this paper the Gaussian marginal copula regression applied to currency exchange rate data set by using log-return transformation and difference transformation according to AIC which are computed by applying the Gaussian marginal copula regression function in GCMR R package.
Keywords: Copula regression, Gaussian copula, Time series

1 Introduction

Various authors discussed likelihood inference for Gaussian copula models. [4] developed a Markov Chain Monte Carlo algorithm for Bayesian inference, [2] adopted a sequential importance sampling algorithm. Well-known limits of the Gaussian copula approach are the impossibility to deal with asymmetric dependence and the lack of tail dependence. These limits may impact the use of Gaussian copulas to model forms of dependence arising, for example, in extreme environmental events or in financial data. Conversely, [5] focuses on working Gaussian copulas used to conveniently handle dependence in regression analysis as described in [3] and [2]. In other terms, the parameters of interest are the regression coefficients, while the dependence structure identified by the Gaussian copula is a nuisance component. In [6] solve the problem of modeling extreme data with the Gaussian copula marginal regression. The model was applied to study the rise harvested area production centers in East Java and other areas. Some useful references can be found in [2] and [3].

Gaussian copula marginal regression models

The known regression model is as it was introduced in [2]:

\[ Y_i = g(x_i, \varepsilon_i; \lambda), \quad i = 1, \ldots, n, \]

where \( g(\cdot) \) is a corresponding function of the regressors \( x_i \) and the error term \( \varepsilon_i \), and \( \lambda \) is a vector of parameters. Among the many possible specifications for the function \( g(\cdot) \), the selection of the model is as follows:

\[ Y_i = F_i^{-1}\{ \Phi(\varepsilon_i); \lambda \}, \]

where \( F_i(\cdot;\lambda) = F(\cdot; x_i; \lambda) \) the cumulative distribution functions of \( Y_i \) given \( x_i \) and \( \Phi(\cdot) \) is standard normal variate. Equation (2) includes all possible parametric regression models for continuous and non-continuous responses. The Gaussian linear regression model \( Y_i = x_i^T \beta + \sigma \varepsilon_i \) corresponds to set \( F_i(Y_i; \lambda) = \Phi\{ (Y_i - x_i^T \beta) / \sigma \} \) in Eq. (2) with \( \lambda = (\beta^T, \sigma)^T \). Model specification is then completed by assuming that the vector of errors \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \) is multivariate normal,

\[ \varepsilon \sim N (0, \Omega), \]

where \( \Omega \) is a correlation matrix.

In Gaussian copula regression the dependence between the variables is modelled with the Gaussian copula so that the joint data cumulative distribution function is given by

\[ P_r(Y_1 \leq y_1, \ldots, Y_n \leq y_n) = \Phi_n(\varepsilon_1, \ldots, \varepsilon_n; P), \]

where \( \varepsilon_i = \Phi^{-1}\{ F(y_i | x_i) \}, \) with \( \Phi_n(\cdot; P) \) the n-dimensional multivariate standard
normal cumulative distribution function with correlation matrix P. This correlation matrix allows to deal with time series, clustered and longitudinal data. In this paper, the Gaussian copula marginal regression (GCMR) in R package will be used. This package gives implementation of Gaussian copula correlation matrices and it makes the forms of dependence arising handled easily, for example, time series and geometrics. The models can be fitting with the method of maximum likelihood inference based on a variate of the Geweke–Hajivassiliou–Keane (GHK) estimator in the continuous case. It also give a select models using information criteria like AIC. In addition, the log likelihood for inference can be computed and the implementation of residuals analysis to evaluate departures from the model assumptions (see [5]).

**Autoregressive Function**

The first step in time series analysis is to determine the form of the model which describes the time series behaviors. This step called “identification”. The basic tool to determine the model is Autoregressive Function (ACF).

Autoregressive Function is a functional relationship between the autocorrelation $\rho_k$ between the values of the time series and the time lags $k$. It has been defined in [7] as

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{cor}(y_t, y_{t-k})}{\sqrt{\text{var}(y_t) \text{var}(y_{t-k})}}.$$  

**Partial Autocorrelation Function**

The Partial Autocorrelation Function (PACF) is the correlation between $y_t$ and $y_{t-k}$ after omitting the effect of $y_t, \ldots, y_{t-k}$ on this correlation.

$$\phi_{kk} = \text{cor}(y_t, \ldots, y_{t-k}|y_{t-k+1}, \ldots, y_{t-1}).$$

PACF is a functional relationship between $\phi_{kk}$ and the lags $k$.

**Autoregressive Moving Average**

Autoregressive Moving Average (ARMA) models are mathematical models of the persistence, or autocorrelation, in a time series. ARMA models are widely used in hydrology, dendrochronology, econometrics, and other fields. Modeling can contribute to understanding the physical system by revealing something about the physical process that builds persistence into the series. In [7] introduced the general ARMA(p, q) process as following:  

$\{X_t\}$ is an ARMA (p,q) process if $\{X_t\}$ is stationary and if for every $t$,  

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{p-1} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$  

where $Z_t$ referred to a white noise (with mean 0 and variance $\sigma^2$), $\{Z_t\} \sim N(0, \sigma^2)$, and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 - \theta_1 z + \cdots + \theta_q z^q)$ have no common factors; $p$ is the number of the Autoregressive model.
(AR) parameters and q is the number of the Moving Average model (MA) parameters.

2 Gaussian copula regression application on currency exchange rate

This section introduces the application for copula by using the Gaussian copula regression with currency exchange rate of Japanese Yen against American dollar denoted by (JPY/USD) and Korean Won against American dollar denoted by (KRW/USD) in the period 1/1/2010 to 31/8/2015, which was taken from the University of British Columbia, Sauder school of business pacific exchange rate service.

First of all we plot the time series for both JPY/USD and KRW/USD separately to detect seasonality, trend and stationarity for both of the time series. It is important to multiply JPY/USD by 10 in order to compare it with KRW/USD.

One can see from Figure 1 that the time series plot could have a seasonality effect, so it cannot be described by using one of the additive model for time series Autoregressive model (AR), Moving Average model (MA), Autoregressive Moving Average model (ARMA) and Autoregressive integrated Moving Average model (ARIMA) until the seasonality is removed.

Now, to study the seasonality which is one of the important component of any time series it should decompose the time series to separate it to its three component which are trend, seasonal and an irregular.
Figure 2a and Figure 2b above show the original time series on the top, the estimated trend component on second from top, the estimated seasonal component on third from top and the estimated irregular component at the bottom. It can be seen that there is a seasonality and the estimated trend component for JYP/USD shows a decrease in 2012 followed by increase from then on to 2015. While in KWR/USD shows a decrease in 2012 and in 2014. Because the JYP/USD and KWR/USD have a seasonality it must be adjust by estimating the seasonal component and subtract the estimated seasonal component from the original time series. By using “decompose ()” function in R package. After the calculation it shows that there is no difference at the time series.

After that, the stationarity should be tested. One way of doing this is by applying the Dickey-Fuller test. The null hypothesis is that there is a unit root, \( \delta = 0 \), which means the differences, \( y_{t+1} - y_t \), are stationary. From Table (1) one can see that neither of these time series is significant at the 5% level \( (p > 0.05) \). So, the data non-stationarity.

<table>
<thead>
<tr>
<th>The time series</th>
<th>Test statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPY/USD</td>
<td>-1.7023</td>
<td>0.6961</td>
</tr>
<tr>
<td>KRW/USD</td>
<td>-1.3619</td>
<td>0.8343</td>
</tr>
</tbody>
</table>
Therefore, it is desirable to have a transformation in order to get a stationarity times. Time series transformation give the benefit to be described using additive model. Log-return, natural log and the difference are the most famous time series transformations.

**Log return transformation**

Let \( x_t \) and \( x_{t-1} \) are two consecutive observations in an exchange rate at time \( t \) and \( t-1 \), and \( y_t \) is the log return (the continuously compounded return) is defined as:

\[
y_t = \log(x_t) - \log(x_{t-1}).
\]  

The benefit of using log returns of series epically financial series is that one can see the changes in the variable.

![Figure 3 Time series of JPY/USD and KWR/USD after the logreturn transformation](image)

From Figure 3 it appears that the time series is stationarity in mean and variance. After satisfying the stationarity model it is possible to use ARMA model as additive model. Since ARMA model include AR and MA parts, it has exactly the same condition of AR(\( p \)) and MA(\( q \)). To select the ARMA model, which means finding the values of the most suitable values of \( p \) and \( q \). It must be explore the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).
Gaussian copula regression application

From Figure 4a and Figure 4b it is visible from the ACF that the autocorrelation at lag 2 exceeds the significance bounds unlike the others do not pass the significance bound. And the PACF shows that the partial autocorrelation at lag 1 pass the significance bounds but the other autocorrelations do not overreach the significance bound. Depending on the ACF and PACF lags the possible ARMA models for the time series will be one of the following: ARMA(0,1), ARMA(1,0), ARMA(2,0), ARMA(2,1), ARMA (0,2) and ARMA (1,2). In Table 2 the different ARMA model are calculated to decide which mode is the best based on AIC in

$$AIC = -2\text{loglik}(\hat{\theta}) + 2p \quad (5)$$
with Gaussian copula marginal regression function gcmr() in R package to study the relationship between JPY/USD and KWR/USD.

Table 2 ARMA models with the currency exchange rate after log return transformation

<table>
<thead>
<tr>
<th>ARMA</th>
<th>AR(1)</th>
<th>AR(2)</th>
<th>MA(1)</th>
<th>MA(2)</th>
<th>AIC</th>
<th>GCMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(0,1)</td>
<td>-</td>
<td>-</td>
<td>0.5075</td>
<td>-</td>
<td>-348.15</td>
<td>-0.0676</td>
</tr>
<tr>
<td>ARMA(1,0)</td>
<td>0.3081</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-343.47</td>
<td>-0.0152</td>
</tr>
<tr>
<td>ARMA(2,0)</td>
<td>0.3792</td>
<td>-0.2458</td>
<td>-0.1089</td>
<td>0.5079</td>
<td>-345.4</td>
<td>-0.0120</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>-0.0669</td>
<td>-0.1089</td>
<td>0.5079</td>
<td>-</td>
<td>-345.16</td>
<td>-0.0595</td>
</tr>
<tr>
<td>ARMA(0,2)</td>
<td>-</td>
<td>-</td>
<td>0.4355</td>
<td>-0.1037</td>
<td>-346.9</td>
<td>-0.0644</td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>-0.1566</td>
<td>0.6098</td>
<td>-</td>
<td>-</td>
<td>-346.74</td>
<td>-0.0677</td>
</tr>
</tbody>
</table>

From Table 2 it is obvious based on AIC the model ARMA(0,1) is the best mode. The Gaussian copula marginal regression made the calculation easier than using the ARMA() function.

After that, it is necessary to study the residuals of the GCMR model because it is very important to find autocorrelation in the residuals and if there is some information not accounted by the model.

Figure 5 The quantile residual for the ARMA(0,1) model

It is clear from Figure 5 the residuals are from a normal distribution and the points tend to fall along the reference line. The ACF plot of the calculated residual shows significant projection at lag 2 in all the different ARMA models. In order to test the autocorrelation, heteroscedasticity and normality three tests are applied. These tests are Box-Pierce test, Ljung-Box test and jarque bera test. The p-value of each test is computed.
Table 3 Residual study for the Gaussian copula regression model with log-return transformation

<table>
<thead>
<tr>
<th>Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box-Pierce</td>
<td>0.6494</td>
</tr>
<tr>
<td>Ljung-Box</td>
<td>0.6421</td>
</tr>
<tr>
<td>jarque bera</td>
<td>LM p-value: 0.513</td>
</tr>
</tbody>
</table>

Table 3 reveals that all the three tests are significant, p-value at lag 1 > 0.05. So that, there is a no serial correlation, no heteroscedasticity and the residual is normal. So, the additive model ARMA are a good model for the data.

**Difference transformation**

Difference transformation attempt to eliminate the trend term by differencing. One can define the first difference operator $\nabla$ by:

$$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t,$$

where B is called the backward shift operator, $B X_t = X_t - 1$. This process should be applied repeatedly until the stationarity is obtained.

Figure 6 shows the time series after the $\nabla = 1$ and it is clear that the stationarity is satisfied.

![Figure 6 Monthly exchange rate JPY/USD and KRW/USD using the difference transformation](image)

From Figure 6 it is clear that the times series are stationary, and one of the following additive model ARMA and ARIMA can be used.
From Figure 7a it is visible that ACF at lag 1 is positive and exceeds the significant bound, while the rest of the ACF will be zero after lag 1 but the PACF in Fig. 6a shows all the lags are significant. In Figure 7b it is clear that ACF at lag 2 is positive and exceeds the significant bound, while the rest of the ACF will be zero after lag 2; but in PACF plot shows lag 1 exceed the significant bound and the rest of the PACF will be zero after lag 1. The possible models will be ARMA(1,0), ARMA(0,1), ARMA(0,2), ARMA(2,0), ARMA(2,1) and ARMA(1,2). After using gcmr() function on the data after the difference of JPY/USD versus KRW/USD.
Table 4 ARMA models with the currency exchange rate after using difference transformation

<table>
<thead>
<tr>
<th>ARMA</th>
<th>AR(1)</th>
<th>AR(2)</th>
<th>MA(1)</th>
<th>MA(2)</th>
<th>AIC</th>
<th>GCMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(0,1)</td>
<td>-</td>
<td>-</td>
<td>0.1768</td>
<td></td>
<td>604.2</td>
<td>0.0436</td>
</tr>
<tr>
<td>ARMA(1,0)</td>
<td>0.2341</td>
<td>-</td>
<td></td>
<td>-</td>
<td>603.25</td>
<td>0.0256</td>
</tr>
<tr>
<td>ARMA(2,0)</td>
<td>0.1950</td>
<td>0.1683</td>
<td>-</td>
<td>-</td>
<td>603.3</td>
<td>0.0202</td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>0.4309</td>
<td>0.1202</td>
<td>-0.2475</td>
<td>-</td>
<td>604.96</td>
<td>0.0223</td>
</tr>
<tr>
<td>ARMA(0,2)</td>
<td>-</td>
<td>-</td>
<td>0.1602</td>
<td>0.1483</td>
<td>604.56</td>
<td>0.0341</td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>0.6728</td>
<td>-</td>
<td>-0.4561</td>
<td>-</td>
<td>603.41</td>
<td>0.0234</td>
</tr>
</tbody>
</table>

From Table 4 depending on AIC value, one can tell that the ARMA(1,0) model is the best fit. Because the data at hand is time series, it is necessary to study the ACF and PACF plot of the residuals. It can reveal if there is any autocorrelation in the residuals.

Figure 8 The quantile residual for the ARMA(1,0) model for the currency data after difference transformation

It can be seen from Figure 8 that there is no departure from reference line so it is easy to say that, the residuals are from a normal distribution and ACF plot explain the remaining residuals show that the model has captured the patterns in the currency exchange rate data after the difference transformation very well, although there is a small amount of autocorrelation left in the residuals the significant spike at lag 2. This suggests that the model can be slightly improved by using any other finical transformation models for time series like autoregressive conditional heteroscedasticity (ARCH) or Generalized autoregressive conditional heteroscedasticity (GARCH) model which is famous in financial time series.
Table 5 Residual study for the Gaussian copula regression model with log-return transformation

<table>
<thead>
<tr>
<th>Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box-Pierce</td>
<td>0.754</td>
</tr>
<tr>
<td>Ljung-Box</td>
<td>0.7487</td>
</tr>
<tr>
<td>Jarque bera</td>
<td>LM p-value: 0.700</td>
</tr>
</tbody>
</table>

Table 5 reveals that the tests for serial correlation, heteroscedasticity are significant. (p-value at lag 1 > 0.05). So, there is no serial correlation, no heteroscedasticity and the residuals are not normal, it is possible to refer this conclusion to the additive model selection. As a financial data it can be use other models.

3. Conclusions

1. Gaussian marginal copula regression applied to currency exchange rate data set it is found that the log-return transformation is better than difference transformation according to AIC which is computed by applying the Gaussian marginal copula regression function in GCMR R package and according to the analysis of the residual.
2. It was found that the GCMR simplifies the application of the time series.

References


Received: September 6, 2016; Published: October 17, 2016