DYNAMIC ECONOMETRIC MODELS
Vol. 6 – Nicolaus Copernicus University – Toruń – 2004

Krzysztof Jajuga
Wrocław University of Economics

Application of Copula Functions in a Modeling of Relations in Multivariate Financial Time Series

1. Introduction – financial econometrics

There are very many models applied to the analysis of economic variables. This obvious statement is particularly relevant when we enter the area of finance. The development of the models used in financial studies is due to at least three factors, namely:

– The development of computer technology (hardware and software) allowing for fast and successful implementation of quantitative models.
– The increasing availability of financial data.
– The development of financial theory.

A large part of modern quantitative methods applied in financial studies is "data driven", that is the specific characteristics of data sets (e.g. varying volatility) are captured by the models designed to match such characteristics.
The analysis of the models applied to describe financial variables leads us to the conclusion that these models can be divided into three general types:

1. Models developed from some underlying financial theory, these models are considered as the hypotheses and then the data analysis is performed to verify these hypotheses.
2. Models developed in the framework of the stochastic approach to be applied and verified in the analysis of the real world; here the underlying model is assumed and then it is verified by using real data.
3. Models developed in the framework of the data-analytic approach; here the underlying model is not known, it is sought by exploring real data.

In the remaining part we discuss the models developed in the stochastic approach. These models, as a rule, are being assigned to **financial economet-**

rics. These are models developed to analyze financial data, for the simulation, forecasting or decision-making objectives. In addition, financial econometrics tools verify the models developed by financial economics and financial mathematics.

The importance of financial econometrics has been highly evaluated by practitioners. On the other hand, also the scientific community considers this field as one of the leading areas in finance. This was confirmed by awarding the Nobel Prize in Economic Sciences in 2003 to Robert Engle "for methods of analyzing economic time series with time-varying volatility (ARCH)" and to Clive Granger "for methods of analyzing economic time series with common trends (cointegration)".

It is well known that most financial data comes in the form of time series. In this paper we consider the multivariate time series as the general form of financial data. The main reasons of the importance of multivariate data are:

- most financial market participants hold portfolios containing more than one financial instrument; therefore they should perform analysis for all components of a portfolio;
- there are more and more financial instruments where payoffs depend on more than one underlying index, among them there are multi-asset derivative instruments (e.g. multi-asset options); therefore to value them one should use models of underlying vectors of indices;
- risk analysis is strongly based on the issue of correlation between the returns (or prices) of the components of a portfolio; therefore multivariate analysis is an appropriate tool to detect the relations between returns.

There are very many models developed by financial econometrics. The more or less extensive surveys and systematizations are given in: Mills (1999), Tsay (2002), Chan (2002), Brooks (2002).

The quantitative methods, applied for multivariate data, try to capture the important characteristics (parameters) of the data set. Among the parameters usually studied are:

- location parameters (e.g. means);
- scale parameters (e.g. standard deviations);
- dependence parameters (e.g. correlation coefficients).

In the case of multivariate time series, the following types of dependence are analyzed:

- dependence between two different variables for the same (or lagged) time units, measured by covariance (correlation);
- dependence between two different time units for the same variable, measured by autocovariance (autocorrelation).

The analysis of multivariate financial time series has recently faced very substantial development. Most of the methods can be put in the following general framework given in the form of the model:

$$X_t = \mu_t + \Sigma_t^{0.5} Z_t,$$
$$\mu_t = E(X_t | X_{t-1}, ...),$$
$$\Sigma_t = E(X_t X_t^T | X_{t-1}, ...).$$

So in this model the multivariate time series is a function of the conditional mean vector and conditional covariance matrix given the past values of this time series. Of course the particular model for multivariate time series depends on the models of the conditional mean vector and the conditional covariance matrix, for example:

– In the case of the conditional mean vector the so-called VARMA model is used;
– In the case of the conditional covariance matrix the so-called MGARCH model is used – given in the most general form as the VECH model.

One of the main problems that arise in this approach is caused by the fact that when the dimension of a multivariate time series increases, there are very many parameters to estimate (even more than a hundred), so computational and also interpretational issue is of real importance. Therefore the simplifications of the most general model were proposed. The most well known models that address the problem of correlation modeling are: the constant correlation model introduced by Bollerslev (1990) and the dynamic conditional correlation model proposed by Engle (2002).

All mentioned models can be treated as somehow classical, because they are based on the (conditional) covariance matrix. The main weaknesses of this approach result from the following features:

– Using covariance (correlation) as a dependence measure is justifiable only for elliptically symmetric distributions;
– Off-diagonal elements of the covariance matrix "integrate" scale and dependence parameters, therefore the measurement of relations may be biased.

The alternative approach to the analysis of the dependence is the so-called copula analysis. We present this approach in the next section.

2. Copula analysis – brief presentation

The key feature of copula analysis is that it gives the decomposition of the multivariate distribution into two components. The first component is marginal distributions. The second component – called the copula function – is the function linking these marginal distributions to form a multivariate distribution. The copula function reflects the structure of the dependence between the components of the multivariate random vector. Therefore the analysis of a multivariate distribution function can be performed by „separating" univariate distributions

from the relation between these distributions. Here the dependence parameters and scale parameters are "separated".
This idea is reflected in the following formula (Sklar (1959)):

$$F(x_1,...,x_n) = C(F_1(x_1),...,F_n(x_n)),$$

where:
F – the multivariate distribution function;
F_i – the distribution function of the ith marginal distribution;
C – copula function.
Therefore, copula function is the distribution function of the multivariate uniform distribution.
The use of copula analysis for the modeling of the relations is suitable because of some properties of the copula function. The most important are the following properties:
– for independent variables we have:

$$C(u_1,...,u_n) = C^{\neg}(u_1,...,u_n) = u_1 u_2 ... u_n,$$

– the lower limit for the copula function is:

$$C^{-}(u_1,...,u_n) = \max(u_1 + ... + u_n - n + 1; 0),$$

– the upper limit for the copula function is:

$$C^{+}(u_1,...,u_n) = \min(u_1,...,u_n).$$

The lower and upper limits for the copula function have important consequences for the modeling of the dependence. To illustrate this problem, suppose that we are given two random variables, X and Y. Two important situations can be distinguished:
– we speak about the total positive dependence between X and Y, when $Y=T(X)$ and T is the increasing function;
– we speak about the total negative dependence between X and Y, when $Y=T(X)$ and T is the decreasing function;

Then it can be proved that:
– in the case of total positive dependence the following relation holds:

$$C(u_1,u_2) = C^{+}(u_1,u_2) = \min(u_1,u_2),$$

– in the case of total negative dependence the following relation holds:

$$C(u_1,u_2) = C^{-}(u_1,u_2) = \max(u_1 + u_2 - 1; 0),$$

Then the set of multivariate distributions with the same marginal distributions can be ordered with respect to the dependence – measured through the copula function. We can describe this by the following formula:

$$C_1(u_1,...,u_n) \le C_2(u_1,...,u_n) \Rightarrow C_1 \prec C_2,$$

and then we have:

$$C^- \prec C^{\neg} \prec C^+.$$

Of course, there are very many possible copula functions. We present here only several copula functions – in bivariate case. All of them depend only on one parameter. The following copulas are studied often in practice:

1. Normal (Gaussian) copula, where:

$$C(u_1,u_2) = \Phi^2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) =$$
$$= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \right) dxdy,$$

2. *t* (Student) copula, where:

$$C(u_1,u_2) = \int_{-\infty}^{t^{-1}(u_1)} \int_{-\infty}^{t^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{p(1-\rho^2)} \right)^{-(p+2)/2} dxdy,$$

3. Farlie-Gumbel-Morgenstern copula, where:

$$C(u_1,u_2) = u_1 u_2 + \theta u_1 u_2 (1-u_1)(1-u_2),$$
$$\theta \in [-1;1],$$

4. Gumbel copula, where:

$$C(u_1,u_2) = \exp(-((-\ln u_1)^\theta + (-\ln u_2)^\theta)^{1/\theta}),$$
$$\theta \in [1;\infty).$$

All these copulas depend on just one parameter, which can be interpreted as dependence parameter. In case of the Gaussian and *t* copula it is correlation coefficient, but in the other cases the dependence has more general meaning than just linear dependence. It can be understood as any type of dependence described by a monotonic function (increasing or decreasing one).

As it was presented, the copula function is a very natural approach to study the dependence for multivariate distributions. It is also well known that the multivariate distribution is a key concept in the theory of stochastic processes. Therefore one can make an attempt to apply the copula approach in modeling time series. We discuss this by considering two possible types of analysis:
– Analysis of the dependence between variables for a particular time unit (data given as multivariate time series);
– Analysis of the dependence between time units for a particular variable (data given as univariate or multivariate time series, in the former case, the analysis is performed separately for each component variable).

Applying the copula analysis to time series is some kind of integration of two independently derived groups of quantitative methods:
– Econometric methods, based on the concept of stochastic process;
– Statistical methods, based on the concept of statistical distribution.

3. The Copula function in multivariate time series analysis – dependence between time units

Here we consider univariate time series (in the case of multivariate time series each component is analyzed separately) and the interest is on the multivariate distribution, whose components are random variables considered in respective time units. Here the order of variables is of importance and in fact, one is (at least implicitly) interested in the conditional distribution of one variable given the other (the others) variables. Therefore, the copula function should in some way reflect the dynamics of the process. This is an alternative approach to the autocovariance (or autocorrelation) function.

To present the approach, we consider univariate stochastic process:

$$X_1, ..., X_n, ...$$

Suppose that the present time unit is denoted by t, and the past time unit by s, where $s < t$.

The main idea behind the use of copula in modeling relations in univariate time series comes from the important results obtained by Darsow, Nguyen and Olsen (1992). These results can be summarized in the following points:

1. The product of two copula functions is defined as:

$$(C_1 \cdot C_2)(u_1, u_2) = \int_0^1 \partial_2 C_1(u_1, u) \partial_1 C_2(u, u_2) du,$$

where:

∂_1, ∂_2 – first order partial derivatives with respect to the first and the second variable.

2. The Theorem on the Markov process can be stated as follows.

The stochastic process is the Markov process if and only if for all positive integers n and for all $t_1, t_2, ..., t_n$; $t_i < t_{i+1}$, the following is true:

$$C_{t_1, t_2, ... t_n} = C_{t_1, t_2} \cdot C_{t_2, t_3} \cdot ... \cdot C_{t_{n-1}, t_n}.$$

This theorem allows for the representation of Markov process through copula functions. Here in the case of the Markov process the relation between n time units is decomposed as a product of the relations between the sequences of two units.

It is worth to note that in the classical representation the Markov process is explained through the initial marginal distribution and transition probabilities (satisfying the Chapman-Kolmogorov equations). In the copula representation the Markov process is explained through marginal distributions and a set of copula functions satisfying the following property:

$$C_{s,t} = C_{s,v} \cdot C_{v,t}.$$

Here the Markov process (for given marginal distributions) depends solely on two-dimensional copula functions. So the operation on copulas corresponds to the operations on transition probabilities used in Markov processes. This makes it possible to model the conditional distribution of the components of stochastic process.

There are two important copula functions for time series: the Brownian copula and the Ornstein-Uhlenbeck copula. They correspond to the two well-known continuous time stochastic processes.

1. Brownian copula.

It is given as:

$$C_{s,t}(u_1, u_2) = \int_0^{u_1} \left(\Phi\left(\frac{\sqrt{t}\Phi^{-1}(u_2) - \sqrt{s}\Phi^{-1}(u)}{\sqrt{t-s}} \right) \right) du,$$

where:
Φ – cumulative standardized normal distribution function.

The most important properties of the Brownian copula are:
– The Brownian copula is a normal copula with the parameter (given as a correlation coefficient) equal to: $\rho = \sqrt{t-s}$.
– If the marginal distributions are normal distributions, then applying the Brownian copula to these distributions leads to the stochastic process, which is the geometric Brownian motion.

– The conditional distribution is given as:

$$P(U_2 \le u_2 | U_1 = u_1) = \Phi\left(\frac{\sqrt{t}\Phi^{-1}(u_2) - \sqrt{s}\Phi^{-1}(u_1)}{\sqrt{t-s}} \right).$$

2. Ornstein-Uhlenbeck copula.

It is given as:

$$C_{s,t}(u_1, u_2) =$$
$$= \int_0^{u_1} \left(\Phi\left(\frac{h(0,s,t)\Phi^{-1}(u_2) - h(0,s,s)\Phi^{-1}(u)}{h(s,s,t)} \right) \right) du,$$
$$h(t0, s, t) = \sqrt{e^{2a(t-s)} - e^{2a(s-t0)}},$$

where:
Φ – cumulative standardized normal distribution function,
a – parameter of this copula.

The most important properties of the Ornstein-Uhlenbeck copula are:
– The Ornstein-Uhlenbeck copula is a normal copula with the parameter:

$$\rho = e^{-a(t-s)} \sqrt{\frac{1-e^{-2as}}{1-e^{-2at}}}.$$

– If the marginal distributions are normal distributions, then applying the Ornstein-Uhlenbeck copula leads to the stochastic process, which is the Ornstein-Uhlenbeck process.
– The limit function for the Ornstein-Uhlenbeck copula, when the parameter a goes to 0, is the Brownian copula and: $\lim_{a \to 0} h(t0, s, t) = \sqrt{t - t0}$.
– The conditional distribution is given as:

$$P(U_2 \le u_2 | U_1 = u_1) = \Phi\left(\frac{h(0,s,t)\Phi^{-1}(u_2) - h(0,s,s)\Phi^{-1}(u)}{h(s,s,t)} \right).$$

It is worth mentioning that the parameter a – being also the mean-reverting coefficient of the Ornstein-Uhlenbeck process – can be interpreted as the parameter of the dependence between random variables being the components of stochastic process – the larger this coefficient, the less dependence between random variables.

It should be added that while using normal distribution as marginal distribution and applying the above mentioned copulas will lead to well known continuous time stochastic processes, one can use other forms of marginal distributions to get other processes. For example, using the t-distribution and the Brownian copula leads to the so called Student (geometric) Brownian motion. This allows for more flexibility in modeling.

4. Copula function in multivariate time series analysis – dependence between variables

Here we will focus on the dependence between two different variables. This dependence can be studied for the same time unit – it is then called concurrent correlation, or for two different time units – it is then called lead-lag correlation. In any case, the focus is on the multivariate distribution at a particular time unit – this is the conditional multivariate distribution, given the past values of all variables.

Since the variables are not ordered, as in the previous case, we can use the regular copula approach. When the copula function is introduced, it replaces the analysis of covariance matrix by the analysis of copula functions. To allow the analysis of multivariate random vector, having some multivariate distribution, by using multivariate time series, one has to assume that univariate stochastic processes are strictly stationary. This is the most natural approach, where one does not pay too much attention to the dynamic properties of univariate time series. These properties are very crucial in classical univariate financial econometrics (for example in the ARIMA-GARCH approach).

However, there is another approach possible, where volatilities and dependences in multivariate time series, both conditional, are modeled separately. This method was presented by Jondeau and Rockinger (2002). The idea is very simple; it combines the univariate time series modeling by the GARCH type models with copula analysis. The easiest implementation of this method is in the case of bivariate time series.

The proposed procedure consists of two steps. In the first step the models for univariate time series are built for both time series. Here the combined procedure of ARIMA models for conditional mean and GARCH models for conditional variance can be used.

In the second step the values of distribution function for residuals obtained after the application of univariate models are subject to copula analysis. Strictly speaking, the dependence parameter for the copula is regarded conditional on the previous values of time series. To make this procedure operational, the unit square of the possible values of two distribution functions is divided into 16 areas (each unit interval for two univariate distribution functions is divided into four subintervals). Finally for each of 16 areas the dependence parameter (assuming a particular copula function) is estimated (for example using the maximum likelihood method). Obviously, this is a conditional dependence parameter. The copula approach in financial econometrics seems to be in its very early stage, however it may produce very interesting results. It is therefore worth exploring.

References

Bollerslev, T. (1990), Modeling the coherence in short-term nominal exchange rates: a multivariate generalized ARCH approach, *Review of Economics and Statistics*, 72, 498–505.

Brooks, C. (2002), *Introductory econometrics for finance*, Cambridge University Press, Cambridge.

Chan, N. H. (2002), *Time series. Applications to finance*, Wiley, New York.

Darsow, W. F., Nguyen, B., Olsen, E. T. (1992), Copulas and Markov processes, Illinois *Journal of Mathematics*, 36, 600–642.

Engle, R. (2002), Dynamic conditional correlation: a simple class of multivariate GARCH models, *Journal of Business and Economic Statistics*, 20, 339–350.

Jondeau, E., Rockinger, M. (2002), Conditional dependency of financial series: the copula-GARCH model, FAME, working paper.

Mills, T. (1999), *The econometric modeling of financial time series*, 2nd ed., Cambridge University Press, Cambridge.

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

Tsay, R. S. (2002), *Analysis of financial time series*, Wiley, New York.