**DYNAMICZNE MODELE EKONOMETRYCZNE**
VIII Ogólnopolskie Seminarium Naukowe, 9-11 września 2003 w Toruniu
Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu

*Krzysztof Jajuga*
*Wrocław University of Economics*

# On modeling the relations in time series

## 1. Introduction – models for financial data

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

– development of computer technology (hardware and software) allowing for fast and successful implementation of quantitative models;
– increase of the availability of financial data, to be applied in the designed models.

The last statement refers particularly to data given as time series, often available as high frequency time series. These data reflect such financial variables as: interest rates, exchange rates, stock prices, market index values, commodity prices. We can draw a conclusion that the 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:

– Models developed from some underlying financial theory, they are considered as the hypotheses and the data analysis is performed to verify these hypotheses;
– Models developed in the framework of stochastic approach to be applied and verified in the analysis of the real world; here the underlying abstract model is assumed and then it is verified through the use of real data;

– Models developed in the framework of 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 classified to the class of financial econometrics. It is worth to draw some distinctions between these models and the models developed within two other areas, namely financial economics and financial mathematics. All three areas contain the models used in financial studies, but there are substantial differences. In our opinion, these areas can be characterized in a following way:

Financial economics – contains the models derived from economic theory, relevant for the analysis of financial processes, particularly these occurring in the financial markets; they are (as a rule) formal mathematical models, based on fundamental economic categories as equilibrium, rational behavior, expectations, etc. It is worth to mention that financial economics is basically the heart of modern theory of finance.

Good example of the model developed in financial economics is Capital Asset Pricing Model (CAPM), drawn from the idea of general equilibrium theory (Sharpe (1964), Lintner (1965)).

Financial mathematics – contains pure mathematical models, based on the concepts, which are abstract and not referred to the theory of finance; these models can be usually interpreted on the ground of the theory of finance and applied in financial studies, this however requires "filling" abstract concepts by financial notions and implementation by financial data analysis.

Good example of the model developed in financial mathematics is martingale pricing model used today for the valuation of derivative instruments (Harrison, Kreps (1979)). This model is commonly interpreted in finance as a risk-free arbitrage pricing model.

Financial econometrics – contains the models developed to analyze financial data, as a rule, given as financial time series, for the simulation, forecasting or decision-making objectives; in addition verifies the models developed by financial economics and financial mathematics.

Good example of the model developed in financial econometrics is GARCH model (Bollerslev (1986)). On the other hand, as the example of the second group of activities in financial econometrics it is worth to mention the procedures to verify models developed within financial economics (Cuthbertson (1996)).

In our opinion, the excellent description of financial econometrics is given in the following two statements (this description "matches" the one given above):

– Journal of Financial Econometrics, describing the scope of this journal: „Designed to address substantive statistical issues raised by the tremendous growth of the financial industry. Papers providing or applying new econometric techniques which are particularly well suited to deal with financial data and models fall within the scope of this journal";

– Monograph on financial econometrics (Campbell, Lo, MacKinlay (1997)): “Raison d’etre of financial econometrics is the empirical implementation and evaluation of financial models”.

## 2. On the models for univariate and multivariate time series

It is well known that the most financial data comes in the form of time series. In this paper we consider the multivariate time series as relatively general form of financial data. There are very many models developed to analyze this type of data. These models were developed by financial econometrics or are simply general econometric and statistical methods applied for financial time series. There are several monographs where the more or less extensive survey and systematization are given, for example: Mills (1999), Tsay (2002), Chan (2002), Brooks (2002).

Of course, we mention here only the methods, by the researchers being constantly classified as financial econometrics methods. In our opinion, from the practical point of view, the suitable classification of these models can be based on two criteria:

– whether the analyzed time series is univariate or multivariate;
– whether the analyzed variables can or cannot be divided into two groups: endogenous variables and exogenous variables.

This second criterion leads to the natural distinction between “regression type models” and “non-regression type models”

Therefore we can distinguish four groups of models:

– univariate time series models – one “endogenous” variable;
– multivariate time series models – many “endogenous” variables;
– multivariate time series models – one endogenous variable, at least one (usually more) exogenous variables;
– multivariate time series models – many endogenous variables, at least one (usually more) exogenous variable.

Now we give very brief remarks on the models belonging to these groups.

### A. Univariate time series models

By no doubt, this is the oldest and the most well known group of models. It is not easy to provide systematic classification of these models. It seems that the substantial part of them can be put in the following framework, based the general form (Tsay (2002)):

$$X_t = g(F_{t-1}) + \sqrt{h(F_{t-1})}\varepsilon_t$$
$$g(F_{t-1}) = \mu_t = E(X_t \mid X_{t-1},....)$$
$$h(F_{t-1}) = \sigma_t^{\ 2} = V(X_t \mid X_{t-1},....)$$

So the basic model contains the conditional mean part and the conditional variance (strictly speaking: standard deviation) part. If we take into account that each of two parts can be linear or nonlinear model, we have the following groups of models:

1. Linear mean and linear variance models – this group contains classical models such as: stationary time series models (ARMA), nonstationary integrated time series models (ARIMA), seasonal models (SARIMA), nonstationary fractionally integrated models (ARFIMA), trend nonstationary models, etc.
2. Linear mean and nonlinear variance models – this group contains classical models mentioned above where additionally variance part is separately modeled by such models as: ARCH, GARCH, SV, CHARMA, GARCH-M, etc.
3. Nonlinear mean and nonlinear variance models – this group contains such models as for example: TAR, STAR, SETAR, bilinear models, Hamilton switching autoregressive models, etc.

B. Multivariate time series models – many “endogenous” variables

Here we have no distinction between endogenous and exogenous variables, all variables are equally treated. Basically there are two groups of models. First group contains the models being the generalization of the models for univariate time series presented above. The most commonly used models are:

– VARMA models, and the special cases: VAR and VMA models;
– Multivariate GARCH models (taken together with models for mean vector, like VAR models, to model mean vector and covariance matrix).

The second type of model within this group comes from the approach based on well known cointegration, which cannot be considered as generalization of univariate models.

C. Multivariate time series models, endogenous and exogenous variables

Basically, these models are the examples of the integration of the theory of stochastic processes with structural econometric modeling. The most general description is within DSEM (Dynamic Structural Equations Model) framework, where the present values of each endogenous variable is described as (linear) functions of past values of this variable, present and past values of other endogenous variables and present and past values of exogenous variables. Among the most important models are those based on the idea of cointegration.

## 3. Analysis of relations – application of copula approach

The quantitative methods, including financial econometrics methods, try to capture the important characteristics (parameters) of the data set. Among the parameters usually studied are:
– location parameters (for example means);
– scale parameters (for example standard deviations);
– relation parameters (for example correlation coefficients).

The careful review of the discussed time series methods indicates that these parameters are key concepts studied in these models. In particular one analyzes the relations:
– between two different variables for the same (or lagged) time units, measured by the covariance (correlation);
– between two different time units for the same variable, measured by the autocovariance (autocorrelation).

This is of course, classical approach. As any method, it has the strengths and weaknesses. The latter results from the following features:
– Using covariance matrix is justifiable only for elliptically symmetric distributions;
– Off-diagonal elements of the covariance matrix "integrate" scale and relation parameters, therefore the measurement of relation may be biased.

The alternative approach to analyze the relations, is based on the so-called copula analysis. We start with the short presentation of this approach, referring to multivariate distribution.

The key point of copula analysis is that it gives the decomposition of the multivariate distribution into two components. The first component consists of the marginal distributions. The second component – called copula function – is the function linking these marginal distributions in multivariate distribution. The copula function reflects the structure of the relation between the components of the multivariate random vector. Therefore the analysis of multivariate distribution function can be performed by „separating" univariate distributions from the relation between these distributions. Thus the relation 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 i-th marginal distribution;
*C* – copula function.

The use of copula analysis for the modeling of the relations is suitable due to some properties of 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 copula function is:

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

– the upper limit for 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 relations. Given two variables, *X* and *Y*, and the function linking these two variables. Two important situations can be distinguished:

– total positive relation between *X* and *Y*, when *Y=T(X)* and *T* is the increasing function;
– total negative relation between *X* and *Y*, when *Y=T(X)* and *T* is the decreasing function.

Then:

– in the case of total positive relationship 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 relationship the following relation holds:

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

The introduction of the copula function leads to the natural order of the multivariate distributions with respect to the relation. 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:

$$C^- \prec C^\neg \prec C^+$$

There are very many possible copula functions studied in the literature. One of the most popular is normal copula, given (in two-dimensional case) as:

$$C(u_1,u_2) = \Phi^2(\Phi^{-1}(u_1),\Phi^{-1}(u_2)) =$$

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

This copula depends on one parameter, $\rho$, which corresponds to correlation coefficient. It is worth to mention that applying normal copula to the marginal distributions other than normal one, leads to other bivariate (in general: multivariate) distributions.

## 4. Copula function in multivariate time series analysis

As it was presented, copula function is very natural (although sometimes difficult to implement) approach to study the relation (dependence) for multivariate distributions. Since the multivariate distribution is key concept in the theory of both univariate and multivariate stochastic processes, it is tempting to adapt copula approach in modeling time series. We discuss this by considering two possible ways:

– studies of the relations between variables for the chosen time unit (data given as multivariate time series);
– studies of the relations between time units for the chosen variable (data given as univariate or multivariate time series, in the former case, the analysis is performed separately for each component variable).

Such an approach is one of the very few ways in which one can “integrate” two independently derived groups of methods:

– econometric methods, based on the concept of stochastic process;
– statistical methods, based on the concept of statistical distribution.

A. Relations between time units

Here we consider univariate time series and the interest is on the multivariate distribution, where components are random variables 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.
To discuss this, we consider univariate stochastic process:

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

and suppose that present time unit is denoted by *t*, and 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). They can be summarized in the following points:

1. Introduction of the product of two copula functions:

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

where:
$\partial_1, \partial_2$ - partial derivative with respect to the first and the second variable.

2. Theorem on Markov process:

The stochastic process is Markov process iff for all positive integers *n* and for all

$$t_1,t_2,...,t_n; \quad t_i < t_{i+1}$$

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 Markov process the relation between *n* time units is decomposed as product of the relations between sequences of two units. In the classical representation Markov process is given through the initial marginal distribution and transition probabilities (satisfying Chapman-Kolmogorov equations). In copula representation Markov process is given through marginal distributions and set of copula functions satisfying the following property:

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

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

There are two important and often discussed copula functions for time series: Brownian copula and Ornstein-Uhlenbeck copula. They correspond to the two continuous time stochastic processes, known under the same names.

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:
$\Phi$ – cumulative standardized normal distribution function.

The most important properties of Brownian copula are:

– Brownian copula is normal copula with parameter:

$$\rho = \sqrt{t-s}$$

– If the marginal distributions are normal distributions, then applying Brownian copula leads to the stochastic process, which is geometric Brownian motion.
– The conditional distribution in the case of Brownian copula 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:
$\Phi$ – cumulative standardized normal distribution function,
$a$ – parameter of this copula.

The most important properties of Ornstein-Uhlenbeck copula are:

– Ornstein-Uhlenbeck copula is normal copula with parameter:

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

– If the marginal distributions are normal distributions, then applying Ornstein-Uhlenbeck copula leads to the stochastic process, which is Ornstein-Uhlenbeck process.
– The limit function for Ornstein-Uhlenbeck copula, when the parameter $a$ goes to 0 is Brownian copula and:

$$\lim_{a\to 0}h(t0,s,t)=\sqrt{t-t0}$$

– The conditional distribution in the case of Ornstein-Uhlenbeck copula is given as:

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

– The parameter $a$ – being also the mean-reverting coefficient of 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 mentioned, that while using as marginal distributions normal distribution and adopting two above mentioned copulas leads to well known continuous time stochastic processes, one can use other forms of marginal distributions to get other processes. For example, using t-distribution and Brownian copula leads to the so called Student (geometric) Brownian motion. This allows for more flexibility in modeling.

B. Relations between variables

Here we consider multivariate time series, but in fact, the interest is on the multivariate distribution in the present moment. Therefore, one is interested in conditional multivariate distribution, given the past information (past values of all variables). Since the variables are not ordered, as in previous case, we can use regular copula approach. When copula function is introduced, it replaces the analysis of covariance matrix by the analysis of copula functions. In practice of course, the problem remains of the choice of suitable copula function. In addition, it is necessary to fulfill the assumption of strict stationarity of stochastic processes.

Despite the promising properties and very appealing interpretation of copulas one has to mention that a lot of work is still to be done, particularly in terms of practical application of these tools.

## References

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, pp. 307-327.

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

Campbell, J.Y., Lo, A.W., MacKinlay, A.C (1997), The econometrics of financial markets, Princeton University Press, Princeton.

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

Cuthbertson, T. (1996), Quantitative financial economics: stocks, bonds and foreign exchange, Wiley, Chichester.

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

Harrison, J.M., Kreps, D.M. (1979), Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory*, 20, pp. 381-408.

Lintner, J. (1965), Security prices, risk and maximal gains from diversification, *Journal of Finance*, 20, pp. 587-615.

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

Sharpe, W.F. (1964), Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, 19, pp. 425-442.

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

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