ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 216, 2008

Krzysztof Jajuga[*]

MULTIVARIATE DISTRIBUTIONS IN FINANCIAL DATA ANALYSIS –APPLICATIONS IN PORTFOLIO APPROACH

ABSTRACT. The paper gives an attempt to systematize different problems in finance which can be solved by multivariate analysis. First of all, some theoretical remarks on multivariate distributions are given. Then, taxonomy of multivariate financial problems is provided.

Key words: multivariate distributions, copula functions, portfolio theory.

I. INTRODUCTION

Technological development, particularly in the area of computer technology, allows for the wide application of advanced statistical tools in financial research and practice. This especially refers to the analysis of multivariate distributions. Among the standard tools used in financial research are:

- Multivariate stochastic process;
- Multivariate distribution;
- Multivariate data analysis methods (regression, discriminant analysis, etc.).

It is worth to mention that using multivariate stochastic process implies also using the concept of multivariate distribution. It is often so that one assumes that the distribution for each random vector in multivariate stochastic process is multivariate normal distribution or at least elliptically symmetric distribution.

There are four main areas of applications of multivariate distributions in financial research:

- Analysis of financial time series;
- Valuation of financial instruments;
- Market risk analysis;
- Credit risk analysis

[*] Professor, Department of Financial Investments and Risk Management Wrocław University of Economics.

The methods of **analysis of financial time series** have two main objectives:

– verification of the hypotheses derived by economic and financial theory using empirical data;

– exploration of data to find some patterns that can be used in the decision making process.

Here the analysis is performed for conditional distribution given past data. In the multivariate case this approach consists in joint modeling of return (conditional mean vector) and volatility and correlation (conditional covariance matrix). Here the general model, VARIMA – MGARCH model, is given as:

$$\mathbf{X}_t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t^{0.5}\mathbf{Z}_t,$$

$$\boldsymbol{\mu}_t = E(\mathbf{X}_t | \mathbf{X}_{t-1},....),$$

$$\boldsymbol{\Sigma}_t = E(\mathbf{X}_t\mathbf{X}_t^T | \mathbf{X}_{t-1},....),$$

where:

$\mathbf{X}_t$ – value (vector) of the multivariate process in time t;

$\boldsymbol{\mu}_t$ – conditional mean vector;

$\boldsymbol{\Sigma}_t$ – conditional covariance matrix;

$\mathbf{Z}_t$ – random process.

So this multivariate process is a sum of the deterministic part, given as conditional mean vector, and the stochastic part, depending on conditional covariance matrix.

There are the other possible general models, for example multivariate stochastic volatility model (MSV) or multidimensional Brownian motion (stochastic process in continuous time).

The main objective of **valuation** models is the determination of the fair value of financial instrument (asset, institution). It is the price at which this instrument should be traded by rational and well informed agents in the market being in equilibrium. Multivariate distributions play crucial role in the valuation of multi-asset options. These are options where instead of one underlying index there are (at least) several underlying indices. The valuation is performed using standard Black-Scholes-Merton framework (e.g. Black, Scholes (1973), Merton (1973)), but for the multivariate underlying process. In the simplest case multidimensional Brownian motion is taken as underlying stochastic process.

The methods of **risk analysis** are particularly well developed for two types of risk, namely:

– market risk – risk resulting from the changes of prices in the financial market (interest rate risk, exchange rate risk, stock price risk, commodity price risk, real estate price risk);

– credit risk – risk resulting from the possibility that the counterparty will not make contractual payments.

A standard approach used in market risk analysis is based on the multivariate distribution of returns – it was originated in Markowitz portfolio theory. The more advanced approach is dynamic one, where one applies multivariate stochastic process of returns.

In credit risk analysis there are two distinct approaches, depending on the analyzed object:

– credit risk single exposure, where the specialized model is designed for specific institution;

– portfolio of homogeneous exposures, where the general model is designed for whole group of enterprises or households with the different parameters for each analyzed object.

Credit risk analysis is usually based on multivariate distribution of losses from different loans.

II. APPLICATIONS OF MULTIVARIATE DISTRIBUTIONS IN PORTFOLIO THEORY

Portfolio theory was one of the first areas in financial research, where the concept of multivariate distributions has been applied. It was proposed by Harry Markowitz (Markowitz (1952)) and extended by James Tobin (Tobin (1958)). It is still considered as a main breakthrough in modern finance. In this classical approach – as the main theoretical concept – multivariate normal distribution and (later) elliptically symmetric distributions were applied.

Classical approach in portfolio theory is based on considering two-criteria decision problem, namely maximizing the level of return (understood as expected return – expected value of the distribution of returns) and minimizing level of risk (understood as standard deviation of the distribution of returns). Then the random variable is considered, being the linear combination of returns of individual financial instruments (e.g. stocks). This approach leads to the strong dependence of the solution on the correlation of returns. This sometimes is criticized as the approach being limited to the linear dependencies between returns, thus being limited to elliptically symmetric distributions.

Here we present the approach, in which:

– instead of two criterions, the combined risk-return criterion is considered;

– the general notion of dependence, free from the drawbacks of correlation coefficient, is assumed.

For simplicity, we consider here bivariate case (returns on two stocks) and we introduce the following notation:

X – return on the first stock, being the random variable;

Y – return on the second stock, being the random variable;

F – cumulative distribution function of X;

G – cumulative distribution function of Y.

Let us also assume that the distributions of X and Y are continuous.

As the portfolio criterion we propose to consider the following criterion:

$$P(X \leq x, Y \leq y), \tag{1}$$

This means that as risk measure one takes the probability that return of EACH stock is below some given level. Clearly, it means that the higher probability, the higher risk.

The criterion given by (1) is in fact risk-return criterion, understood as the probability of falling (risk driven notion) below some level (return driven notion). Also, this is the criterion which goes beyond Markowitz approach, since:

$$X \leq x \text{ and } Y \leq y \Rightarrow \alpha X + (1-\alpha)Y \leq \alpha x + (1-\alpha)y.$$

But the opposite implication is not true.

The classical analysis of multivariate distribution (including bivariate distribution of returns) is based on the covariance matrix. It is thus assumed that all information about dependence between the components of the random vector is contained in covariance matrix. Since the dependence is the crucial notion in portfolio analysis, it means that in Markowitz approach analysis is covariance-driven (correlation-driven).

We adopt here an alternative approach to analyze the multivariate distribution (of returns), where instead of analyzing jointly individual risk parameters and dependence parameters, given in the covariance matrix, the analysis is performed separately for individual risk parameters (through the analysis of univariate distributions) and for dependence parameters. This approach is based on the so-called copula functions.

The main idea of copula analysis lies in the decomposition of the multivariate distribution into two components. The first component – it is marginal distributions. The second component is the function linking these marginal distributions to get a multivariate distribution. This function reflects the structure of the dependence between the components of the random vector. Therefore the analysis of multivariate distribution function is conducted by „separating" univariate distribution from the dependence.

This idea is reflected in Sklar theorem (Sklar (1959)), given as:

$$H(x_1,...,x_m) = C(H_1(x_1),...,H_m(x_m)), \quad (2)$$

where:

H – the multivariate distribution function;

H_i – the distribution function of the i-th marginal distribution;

C – copula function.

Since we consider here the bivariate distribution (two-stock portfolio), we get the special version of the theorem given by (2):

$$P(X \le x, Y \le y) = H(x, y) = C(F(x), G(y)). \quad (3)$$

So in this case copula function is simply the distribution function of the bivariate uniform distributions. The bivariate distribution function is given as the function of the univariate (marginal) distribution functions. This function is called copula function and it reflects the dependence between the univariate components.

To recapture, we have the following risk equations in both approaches:

– Markowitz approach:

$$V(\alpha X + (1-\alpha)Y) = \alpha V(X) + (1-\alpha)V(Y) + 2\alpha(1-\alpha)COV(X,Y),$$

– Copula approach:

$$P(X \le x, Y \le y) = C(P(X \le x), P(Y \le y)).$$

So the main difference in these two approaches lies in the fact that copula approach considers the returns on the components of the portfolio separately and Markowitz approach considers the return on whole portfolio.

As one can see in copula approach, risk of a two-stock portfolio is a function of the individual risk of each stock (understood as falling below some level of return) and the copula function "linking" these two individual risks. Therefore copula function, reflecting dependence, plays similar role as correlation of returns in Markowitz approach.

There are many possible copula functions which can be applied. Three important copula functions are (given in bivariate case):

– Copula with independent variables:

$$C(F(x),G(y)) = F(x)\cdot G(y), \tag{4}$$

– Lower limit copula:

$$C(F(x),G(y)) = \max(F(x)+G(y)-1;0), \tag{5}$$

– Upper limit copula:

$$C(F(x),G(y)) = \min(F(x),G(y)). \tag{6}$$

Application of these three copula functions leads to three two-stock portfolios, defined in the following way:

– the lowest risk portfolio, obtained for lower limit copula function, given as:

$$P(X \le x, Y \le y) = \max(P(X \le x) + P(Y \le y) - 1;0), \tag{7}$$

– the highest risk portfolio, obtained for upper limit copula function, given as:

$$P(X \le x, Y \le y) = \min(P(X \le x), P(Y \le y)), \tag{8}$$

– portfolio with independent components, given as:

$$P(X \le x, Y \le y) = P(X \le x)\cdot P(Y \le y). \tag{9}$$

The portfolios given by (7), (8) and (9), obtained for copula approach correspond to Markowitz two-stock portfolios, in which correlation of returns is equal to -1, +1 and 0, respectively.

In addition, many other copula functions are considered in the theoretical and empirical studies, for example:

– Gaussian (normal) copula:

$$C(F(x),G(y)) = \int_{-\infty}^{\Phi^{-1}(F(x))}\int_{-\infty}^{\Phi^{-1}(G(y))} \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right)dxdy, \tag{10}$$

– Gumbel copula, given as:

$$C(F(x),G(y)) = \exp(-((-\ln F(x))^{\theta} + (-\ln G(y))^{\theta})^{1/\theta}),$$
$$\theta \in [1;\infty), \tag{11}$$

– Clayton copula, given as:

$$C(F(x),G(y)) = \max((F(x)^{-\theta} + G(y)^{-\theta} - 1)^{-1/\theta};0),$$
$$\theta \in [-1;\infty), \theta \neq 0 \tag{12}$$

– Ali-Mikhail-Haq copula, given as:

$$C(F(x),G(y)) = \frac{F(x) \cdot G(y)}{1-\theta(1-F(x))(1-G(y))},$$
$$\theta \in [-1;1], \tag{13}$$

– Frank copula, given as:

$$C(F(x),G(y)) = -\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta \cdot F(x)}-1)(e^{-\theta \cdot G(y)}-1)}{(e^{-\theta}-1)}\right),$$
$$\theta \neq 0, \tag{14}$$

– Farlie-Gumbel-Morgenstern copula, given as:

$$C(F(x),G(y)) = F(x) \cdot G(y) + \theta \cdot F(x) \cdot G(y) \cdot (1-F(x)) \cdot (1-G(y)),$$
$$\theta \in [-1;1]. \tag{15}$$

Copula functions given by (11)–(15) are one-parameter functions. The parameter θ can be interpreted as the dependence parameter, where the dependence is understood in more general sense than correlation.

It is also worth to mention that classical analysis of bivariate normal distribution can be put in the framework of copula analysis, by assuming univariate normal distribution as marginal distribution and choosing normal copula. Of course, applying other copula functions to normal marginals leads to the distributions other than bivariate normal.

The detailed presentation of different copula functions is given in Nelsen (1999) and Joe (1997).

Table 1 presents simple example of portfolio risk for different copula functions. We consider the case of medians of marginal distributions of returns:

$$P(X \le x) = 0.5; P(Y \le y) = 0.5.$$

As one can see, the proposed approach is more general that classical portfolio approach and it can be suited for the distributions of returns different from elliptically symmetric distributions. From table 1 it is clear that are different limits for portfolio risk in the case of different dependence structures.

Table 1

Risk of two-stock portfolios for different copula functions

Copula	$P(X \leq 0.5, Y \leq 0.5)$
Lower limit	0
Independence	0.25
Upper limit	0.5
Normal, correlation: -0.9	0.072
Normal, correlation: -0.5	0.167
Normal, correlation: 0	0.25
Normal, correlation: 0.5	0.333
Normal, correlation: 0.9	0.428
Farlie-Gumbel-Morgenstern θ=-1	0.1875
Farlie-Gumbel-Morgenstern θ=0	0.25
Farlie-Gumbel-Morgenstern θ=1	0.3125
Ali-Mikhail-Haq θ=-1	0.2
Ali-Mikhail-Haq θ=0	0.25
Ali-Mikhail-Haq θ=0.95	0.3279
Gumbel θ=1	0.25
Gumbel θ=5	0.4510
Clayton θ=-1	0
Clayton θ=1	0.3333
Clayton θ=10	0.4665
Frank θ=-10	0.0686
Frank θ=-1	0.2191
Frank θ=1	0.2809
Frank θ=10	0.3125

REFERENCES

Black F., Scholes M. (1973), *The pricing of options and corporate liabilities*, Journal of Political Economy, 81, pp. 637–654.

Joe H. (1997), *Multivariate models and dependence concepts*, Chapman and Hall, London.

Markowitz H. M. (1952), *Portfolio selection*, Journal of Finance, 7, pp. 77–91.

Merton R.C. (1973), *Theory of rational option pricing*, Bell Journal of Economics and Management Science, 4, pp. 141–183.

Nelsen R.B. (1999), *Introduction to copulas*, Springer, New York.

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.

Tobin J. (1958), *Liquidity preference as behavior towards risk*, Review of Economic Studies, 25, pp. 65–86.

Krzysztof Jajuga

ROZKŁADY WIELOWYMIAROWE W ANALIZIE DANYCH FINANSOWYCH – PRÓBA SYSTEMATYZACJI

W artykule przedstawiona jest próba systematyzacji różnych problemów finansowych, do rozwiązania których stosowane są metody analizy wielowymiarowej. Na wstępie podane są uwagi na temat rozkładów wielowymiarowych. Następnie proponowana jest taksonomia wielowymiarowych problemów finansowych.