Monte Carlo Simulations of the multivariate distributions with different marginals*

Mária Bohdalová
Comenius University, Faculty of Management, Department of Information Systems
e-mail: maria.bohdalova@fm.uniba.sk

Ľudomír Šlahor
Comenius University, Faculty of Management, Department of Finance and Economics
e-mail: ludomir.slahor@fm.uniba.sk

Abstract

In this paper, we will describe Monte Carlo simulations of the multivariate distributions with different marginals. This approach demands the joint distribution to be known (marginals models can be fitted using different distributional specifications, including nonnormal distributions) and it uses copula theory as a fundamental tool in modeling multivariate distribution. The copula theory allows the definition of the joint distribution through the marginal distributions and the dependence between the variables. It is obvious, that the accuracy of the forecasts of the behavior of the risk factors depends on the sources of error. Usually, the residuals follow normal distribution. However, it has been observed that the Gaussian models do not fit very well the real-life data, e. g. do not allow so called fat tails and asymmetry of observed log returns. Fat tail is a property of some probability distributions (alternatively referred to as heavy tail distributions) exhibiting extremely large kurtosis. In this paper, we propose a multivariate simulation using models that have different error distribution.

1 Introduction

Modeling of the correlated random variables requires knowledge of their joint (multivariate) distribution. However, the available real data regarding the association between random variables is often limited to some summary statistics (e.g. correlation matrix). In the special case of multivariate normal distributions, the covariance matrix and the mean vector, as summary statistics, completely specify the joint distribution. Generally, specific dependency models have to be used in conjunction with summary statistics. When we take into account higher dimensions, the variety of dependency structures dramatically increases. For given fixed marginal distributions and correlation matrix, we can construct infinitely many joint distributions.

In practice, a wide array of distributions can be used for different random variables (e.g.risk factors in finance sector). Some of the commonly used distributions are the normal, the lognormal, geometric Brownian motion, GARCH, and others. The key to a meaningful modeling of the random variables is making reliable judgments about which statistical distribution is appropriate for which random variables and estimating the parameters of the selected distributions. Another important issue is the specification of a modeling structure that meaningfully takes into account the interrelationships between different random variables.

*This research was supported by Science and Technology Assistance Agency under contracts No. VEGA-1/3014/06, APVV-1/4024/07 and VEGA-0375-06

Of course, the accuracy of the forecasts of the behavior of the random variables depends on the sources of error of the model. Usually, the residuals follow normal distribution. However, it has been observed that the Gaussian models do not fit very well the real-life data, e. g. do not allow so called fat tails and asymmetry of observed distributions.

In this paper, we propose to use copula function for describing joint distributions given by the marginal distributions of the random variables and we use a multivariate t-distribution as a simple and powerful tool to examine the residuals.

The paper is organized as follows. Section 2 introduces some basics definitions, copula function and presents Sklar's theorem. Monte Carlo simulation with dynamic random vector modeling is presented in Section 3. Finally, section 4 contains an application of the method to a practical example.

2 Some basics definitions and copulas

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be the random vector of the n random variables with marginal cumulative distribution functions (C.D.F.) $F_1, \ldots, F_n$. The multivariate C.D.F. $F(x_1, \ldots, x_n) = P[X_1 \leq x_1, \ldots, X_n \leq x_n]$, completely determines the dependence structure of random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$. However, its analytic representation is often too complex, making practically impossible its estimation and consequently its use in simulation models.

The most common methodologies use the multivariate conditional Gaussian distribution to simulate random vector values due to its easy implementation. Unfortunately, empirical evidence underlines its inadequacy in fitting real data. The use of a copula function allows us to overcome the issue of estimating the multivariate C.D.F. by splitting it into two parts:

- determine the margins $F_1, \ldots, F_n$, representing the distribution of each random variables; estimate their parameters fitting the available data by soundness statistical methods[1];
- determine the dependence structure of the random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$, specifying a meaningful copula function.

Copulas are parametrically specified joint distributions generated from given marginals. Therefore, properties of copulas are analogous to properties of joint distributions ([14], p. 7). The theory of the copulas is a very powerful tool for modeling joint distribution because it does not require the assumption of joint normality and allow us to decompose any n-dimensional joint distribution into its n marginal distributions and a copula function. Conversely, a copula produces a multivariate joint distribution combining the marginal distributions and the dependence between the variables. Copula have been broadly used in statistical literature. The books of [8] and [11] presented a good introduction to the copula theory. The copula theory have been only recently used in the financial area, there are already several applications in this area. The papers of [2], [5] and [4] provided general examples of applications of copula in finance. While the method is borrowed from the theory of statistics, it has been gathering strong popularity both among academics and practitioners in the financial world mainly because of the increase of volatility and erratic behavior of financial markets.

In this section, we will, at first, introduce the general definition of copula and Sklar's Theorem.

Definition 2.1. *An n - dimensional copula[2] is a multivariate C.D.F. C, with uniformly distributed margins on $\langle 0, 1\rangle$ $(Uniform\,(0,1))$ and the following properties:*

1. $C : \langle 0, 1\rangle^n \rightarrow \langle 0, 1\rangle$;

[1] e.g., Generalized Method of Moments (GMM), Maximum Likelihood Estimation (MLE)., etc.

[2] The original definition is given by [13].

2. *C is grounded and n-increasing*[3]*;*

3. *C has margins C_i which satisfy $C_i(u) = C(1, \ldots, 1, u, 1, \ldots, 1) = u$ for all $u \in \langle 0, 1 \rangle$.*

It is obvious, from the above definition, that if $F_1, \ldots, F_n$ are univariate distribution functions, $C(F_1(x_1), \ldots, F_n(x_n))$ is a multivariate C.D.F. with margins $F_1, \ldots, F_n$, since $U_i = F_i(X_i)$, $i = 1, \ldots, n$, is a uniform random variable. Copula functions are a useful tool to construct and simulate multivariate distributions.

The following theorem is known as Sklar's Theorem. It is the most important theorem regarding to copula functions since it is used in many practical applications.

Theorem 2.1. *Let F be an n-dimensional C.D.F. with continuous margins $F_1, \ldots, F_n$. Then F has the following unique copula representation (canonical decomposition):*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{1}$$

The proof can be found in ([11], p. 18).

The theorem of Sklar [13] is very important, because it provides a way to analyse the dependence structure of multivariate distributions without studying marginals distributions. From Sklar's theorem we see, that for continuous multivariate distribution functions, the univariate margins and the multivariate dependence structure can be separated. The dependence structure can be represented by an adequate copula function. Moreover, the following corollary is attained from eq. 1.

Corollary 2.2. *Let F be an n-dimensional C.D.F. with continuous margins $F_1, \ldots, F_n$ and copula C (satisfying eq. 1). Then, for any $u = (u_1, \ldots, u_n)$ in $\langle 0, 1 \rangle^n$:*

$$C(u_1, \ldots, u_n) = F\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right) \tag{2}$$

where F_i^{-1} is the generalized inverse of F_i.

To illustrate the idea behind the copula function, we can think about the multivariate Gaussian distribution that is a "standard" assumption in risk management.

Corollary 2.3. *The Gaussian (or normal) copula is the copula of the multivariate normal distribution. In fact, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is multivariate normal iff:*

1. *the univariate margins $F_1, \ldots, F_n$ are Gaussians;*

2. *the dependence structure among the margins is described by a unique copula function C (the normal copula) such that*[4]*:*

$$C_R^{Ga}(u_1, \ldots, u_n) = \Phi_R\left(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_n)\right), \tag{3}$$

   where Φ_R is the standard multivariate normal C.D.F. with linear correlation matrix R and ϕ^{-1} is the inverse of the standard univariate Gaussian C.D.F.

Although the normal copula does not have a simple analytical expression, it lends itself to simple Monte Carlo simulation techniques.

[3] These properties mean that C is a positive probability measure.

[4] As one can easily deduce from eq. 2.

3 Monte Carlo simulation

Monte Carlo simulation is a general method of modeling stochastic processes (i.e. processes involving human choice or processes for which we have incomplete information). It simulates such a process by means of random numbers drawn from probability distributions which are assumed to accurately describe the uncertain components of the process being modeled ([12], p. 22). Monte Carlo simulation has become a key technology in the financial sector. It can be applied in a variety of settings.

The traditional Monte Carlo simulation method is based on the following.

Definition 3.1. *Assume that random variable X has a cumulative distribution function (C.D.F.) F_X. We define F_X^{-1} as*

$$F_X^{-1}(q) = inf\{x : F_X(x) \geq q\}, \ 0 < q < 1. \tag{4}$$

Remark that F_X^{-1} is non-decreasing.

Lemma 3.1. *For any random variable X and any random variable U which is uniformly distributed on $(0,1)$, we have that X and $F_X^{-1}(U)$ have the same C.D.F.*

For the proof see [15], p. 3.

A Monte Carlo simulation of a random variable X can be achieved by first drawing a random uniform number u from $U \sim Uniform(0,1)$, and then inverting u by $x = F_X^{-1}(u)$.

In a similar way, a Monte Carlo simulation of the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, usually starts with uniform random vector $\boldsymbol{U} = (U_1, \ldots, U_n)$. If the random variables $X_1, \ldots, X_n$ are independent (correlated), then we need n independent (correlated) uniform random variables $U_1, \ldots, U_n$. For a set of given marginals, the correlation structure of the random variables $X_1, \ldots, X_n$ is completely determined by the correlation structure of the uniform random variables $U_1, \ldots, U_n$ (see [15], p. 3).

Monte Carlo simulation of the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ usually assumes that this vector has a multivariate normal distribution with normal marginals $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ (μ_i and σ_i^2 denote the mean and the variance of X_i) and a positive definite covariance matrix Σ:

$$\mu_i = \mathbb{E}(X_i), \ \sigma_i^2 = \mathbb{E}\left((X_i - \mu_i)^2\right), \tag{5}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & c_{1,2} & c_{1,3} & \cdots & c_{1,n} \\ c_{1,2} & \sigma_2^2 & c_{2,3} & \cdots & c_{2,n} \\ c_{1,3} & c_{2,3} & \sigma_3^2 & \cdots & c_{3,n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ c_{1,n} & c_{2,n} & \cdots & c_{n-1,n} & \sigma_n^2 \end{pmatrix}, \tag{6}$$

where $c_{i,j}$ is the covariance between X_i and X_j:

$$c_{i,j} = \mathbb{E}\left((X_i - \mu_i)(X_j - \mu_j)\right). \tag{7}$$

Detailed information about Monte Carlo simulation is e.g. in [8], [9], [7], [3], [6].

In a more complex form, statistical simulation would depend on more complex assumptions than normality and stable correlations. It would incorporate kurtosis and skewness in the probability distributions of changes in random variables, dynamic correlations of changes in random variables (e.g. the correlations could depend on the magnitude of changes of random variables) and fat tails.

In this paper, we describe Monte Carlo simulation with dynamic random vector modeling. Monte Carlo simulation with dynamic random vector modeling is done in the following sequence:

1. Suppose, that is given a random vector $\boldsymbol{X}$ of n correlated random variables $X_1, \ldots, X_n$. Each random variable may be modeled by appropriate model with typical equation:

$$y_i = f_i(\boldsymbol{x}, \boldsymbol{y}, \theta_i) + \epsilon_i \tag{8}$$

where $i = 1, 2, \ldots, n$, $\boldsymbol{y}$ are the endogenous variables, $\boldsymbol{x}$ are the exogenous variables, θ_i are the estimated parameters, and ϵ_i are the residual errors:

$$\epsilon_i \sim F_i(\chi_i), \tag{9}$$

where $F_i(.)$ is a user-specified distribution function. (The estimation and inference of the parameter vectors θ and χ is based on standard maximum likelihood estimation (MLE) procedures and generalized method of moments (GMM) and they can be estimated using desktop econometric software such as SAS® software.)

2. Monte Carlo simulations are based on random draws ϵ from a variable with the desired probability distribution ([9], p.295). Therefore we need to estimate the parameters of the joint distribution $F(.)$ for the error vector, that is determine by the copula $C(.)$ (e.g. Gaussian copula, see eq. 3).

   The estimate of the joint distribution $H(.)$ for the error vector can be constructed as: $F(\epsilon) = C\left(F_1^{-1}(\epsilon_1), \ldots, F_n^{-1}(\epsilon_n)\right)$ (see eq. 2).

   (This fixes the distribution for $\boldsymbol{y}$, the endogenous variables, because the parameters of the model equations have already been estimated (in the last step), and the equations can be used to solve for $\boldsymbol{y}$.)

3.
   - In the case of the multinormal error vector, we estimate correlation matrix Σ using the C.D.F.s ($F_i(.)$), along with the inverse standard normal C.D.F. (ϕ^{-1}), which uses the relationship $\phi^{-1}(F_i(\epsilon_i))$.
   - In the case of the multivatiate t-distribution of error vector, correlation matrix Σ is created by crossing the normally distributed residuals. The normally distributed residuals are created from the t-distribution residuals using the normal inverse C.D.F. and the t-C.D.F.

4. This step consists of a generation m (m is the number of Monte Carlo iterations, usually m equals to 1000, 1500 and more) n-tuples of independent $N(0,1)$ variables, that are transformed to a correlated set by using Σ[5]. They are then transformed back to the uniform by using $\phi^{-1}(.)$. These u_i are transformed into a set of draws from the joint distribution by using $\epsilon_i = \phi_i^{-1}(u_i)$.

In the next section, we will show how this method can be used for modeling of the incoming joint probability density functions in financial sector using multivariate t-distribution and alternatively normal distribution.

4 Example

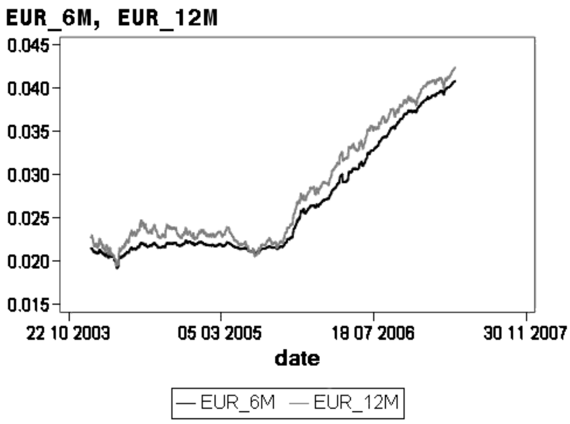
In the financial sector, interest rates are a very important subject. Interest rates, which can be loosely defined as the price of money, affect the livelihoods of individuals and businesses each and every day. In particular, we are interested in the LIBOR rates since they serve as a widely used reference interest rate for a variety of financial instruments. LIBOR stands for the London

[5]The standard techniques for simulating correlated changes in n risk factors are Cholesky decomposition and Principal Component Analysis - PCA. For a description see ([10], p. 80), ([9], p. 303), ([1], p. 41-52)

Interbank Offered Rate and is an interest rate at which banks can borrow funds, in marketable size, from other banks in the London interbank market. LIBOR rates are currently fixed in various currencies (USD, EUR, GBP, JPY and other major currencies) and maturities (usually 1, 3, 6 or 12 months) and is widely used as an interest rate index. LIBOR is the primary benchmark used by banks and investors worldwide. Aside from fixing the cost of borrowing money, LIBOR is used to calculate interest rates applied to a wide range of contracts including over-the-counter (OTC) instruments such as interest rate swaps, forward rate agreements, syndicated loans, floating rate notes, etc. In our investigation we deal exclusively with six and twelve monthly maturities in both USD and EUR rates.

4.1 Data description

The presented theory is applied to a portfolio composed by LIBOR's interest rates. The database contains 841 daily interest rates of EUR 6M, EUR 12M, USD 6M, USD 12M, from January 2nd 2004 to April 11th 2007. Data follows from `http://www.bba.org.uk/bba/jsp`. Figure 1 presents the plots of all series risk factors and Table 1 contains descriptive statistics.



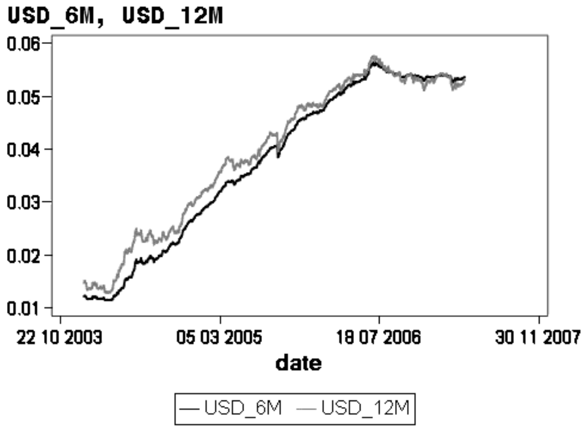


Figure 1: Daily interest rates of EUR & USD 6M and 12M (LIBOR)

From the figure 1, we can see that both, the EUR and USD time series, are correlated. Table 1 shows that means of all series are positive. All time series distributions are nearly symmetric, but they do not have normal distributions (p-value of the Shapiro-Wilk test of normality is less than < 0.0001 for all time series).

Variable	**Mean**	**Std.Dev**	**Min**	**Max**	**Median**	**Skewness**	**Kurtosis**
EUR 6m	0.02654	0.00654	0.01920	0.04086	0.02210	0.88032	-0.76313
EUR 12m	0.02796	0.00686	0.01943	0.04240	0.02383	0.74235	-0.99976
USD 6m	0.03753	0.01512	0.01145	0.05640	0.04000	-0.36145	-1.31994
USD 12m	0.03942	0.01393	0.01283	0.05766	0.04220	-0.46377	-1.16588

Table 1: Descriptive statistics of the risk factors

4.2 The model

In specifying the fourvariate model we must, at first, define the four models for the marginal variables and then the error vector model for copula. The models for the univariate variables must take into account the characteristics of the risk factor. We use multivariate geometric Brownian motion models of the following form:

$$
\begin{aligned}
rate_t &= rate_{t-1} + \mu \cdot rate_{t-1} + \nu \\
\nu &= \sigma \cdot rate_{t-1} \cdot \epsilon \\
\epsilon &\sim N(0,1).
\end{aligned}
\tag{10}
$$

For the purpose of this paper we perform Monte Carlo analysis of models that have residuals distributed as

1. a multivariate t-distribution ($\epsilon \sim t\left(\sigma_1^2, \ldots, \sigma_n^2, df\right)$)
2. a multivariate normal distribution ($\epsilon \sim N\left(0, \Sigma\right)$).

The joint probability density function of the interest rates (EUR 6M and EUR 12M, USD 6M and USD 12M, EUR 6M and USD 6M, EUR 12M and USD 12M) on the fifth day a-head are shown in the figures 2–5 in Appendix.

4.3 Concluding remarks

In this paper, we have shown that the t-distribution with copula theory can be a very powerful tool in forecasting the incoming joint PDF of the risk factors. We used multivariate geometric Brownian motion model with an error vector having t-distribution and alternatively normal distribution. The calculations were obtained with the software system SAS® and SAS® Risk Dimensions.

4.4 Acknowledgments

We greatly acknowledge grants from Tatra Banka and SAS Slovakia.

References

[1] M. Bohdalová and I. Stankovičová. Využitie pca pri analýze rizikových faktorov investičného portfólia. *Forum Statisticum Slovakum*, (3):41–52, 2006.

[2] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli. Copulas for finance, a reading guide and some applications., 2000. Working paper, Financial Econometrics Research Center, City University, London.

[3] M. Crouhy, D. Galai, and R. Mark. *Risk Management*. McGraw-Hill Companies, New York, 2001.

[4] P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. pages 329–384. Handbook of Heavy Tailed Distributions in Finance (Edited by S.T. Rachev), Elsevier, 2003.

[5] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls. pages 176–223. Risk Management Value at Risk and Beyond (Edited by M. Dempster), Cambridge University Press, 2002.

[6] W. Graeme. Risk measurement for financial institutions, 2007. http://www.smealsearch.psu.edu/29776.html.

[7] J. Jílek. *Finanční rizika*. Grada Publishing, Praha, 2000.

[8] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman and Hall, 1997.

[9] P Jorion. *Value at Risk: The Benchmark for Controlling Market Risk.* McGraw-Hill Professional Book Group, Blacklick, USA, 2000.

[10] S. Míka. *Numerické metódy algebry.* SNTL, Praha, 1985.

[11] R. B. Nelsen. *An Introduction to Copulas.* Springer Verlag, New York, 1999.

[12] E. Picoult. *Risk Management: Value at Risk and Beyond*, chapter Quantifying the Risks of Trading, pages 1–59. Cambridge University Press, 2002.

[13] A. Sklar. Random variables, distribution functions and copulas - a personal look backward and forward, in distributions with fixed marginals and related topics. pages 1–14. Institute of Mathematical Statistics, (Edited by L. Rüschendorff, B. Schweizer and M. Taylor), Hayward, CA., 1996.

[14] P. Trivedi and D. Zimmer. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, Vol. 1:1–111, 1 2007.

[15] S. Wang. Aggregation of correlated risk portfolios: Models & algorithms. *Proceedings of the Casualty Actuarial Society*, pages 848–939, 1998.

Mária Bohdalová,
Department of Information systems,
Comenius University, Faculty of Management,
Odbojárov 10,
820 05 Bratislava 25
e:mail: maria.bohdalova@fm.uniba.sk

Ľudomír Šlahor,
Department of Finance and Economics,
Comenius University, Faculty of Management,
Odbojárov 10,
820 05 Bratislava 25
e:mail: ludomir.slahor@fm.uniba.sk

Appendix

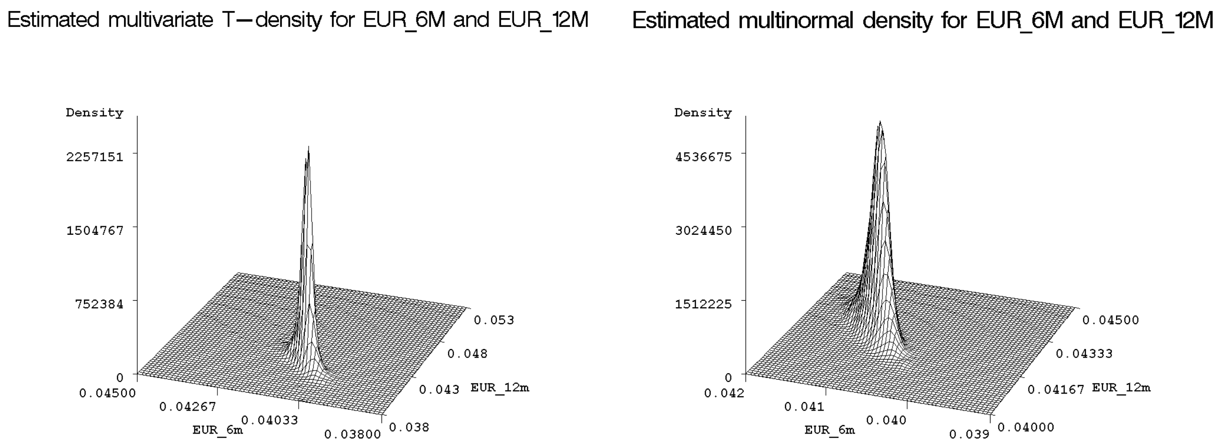


Figure 2: Joint PDF of the EUR 6M and EUR 12M (error vectors have multiv.t- and normal-distribution)

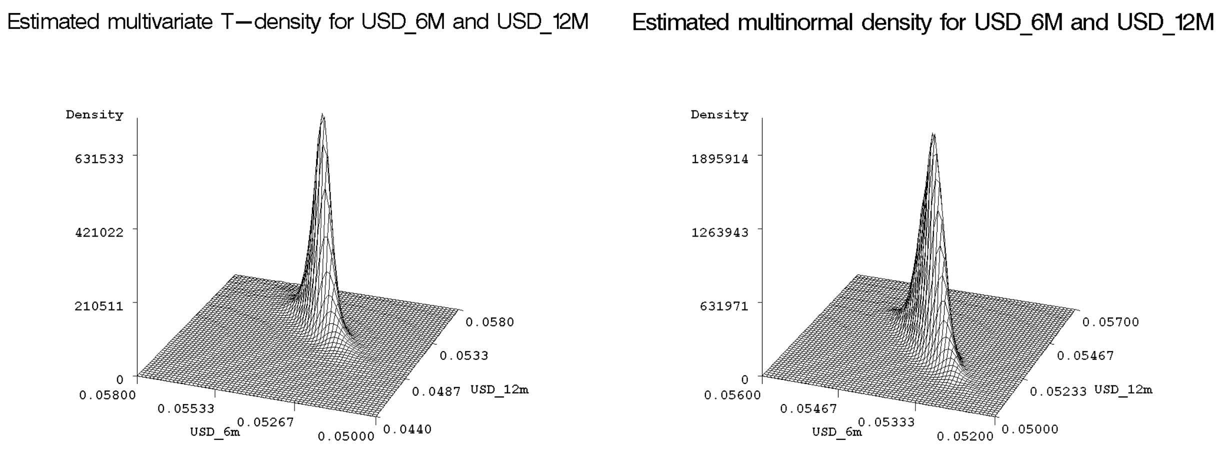


Figure 3: Joint PDF of the USD 6M and USD 12M (error vectors have multiv.t- and normal-distribution)

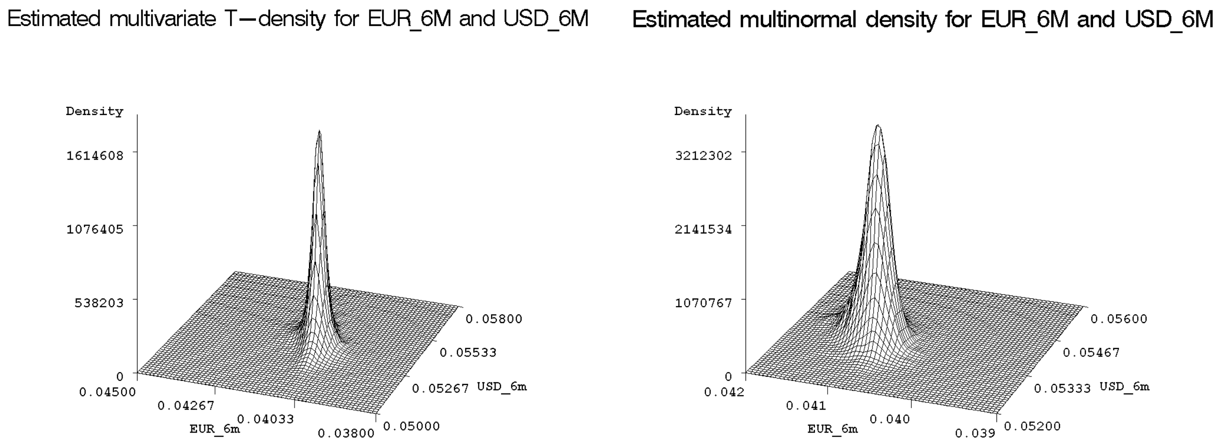


Figure 4: Joint PDF of the EUR 6M and USD 6M (error vectors have multiv.t- and normal-distribution)

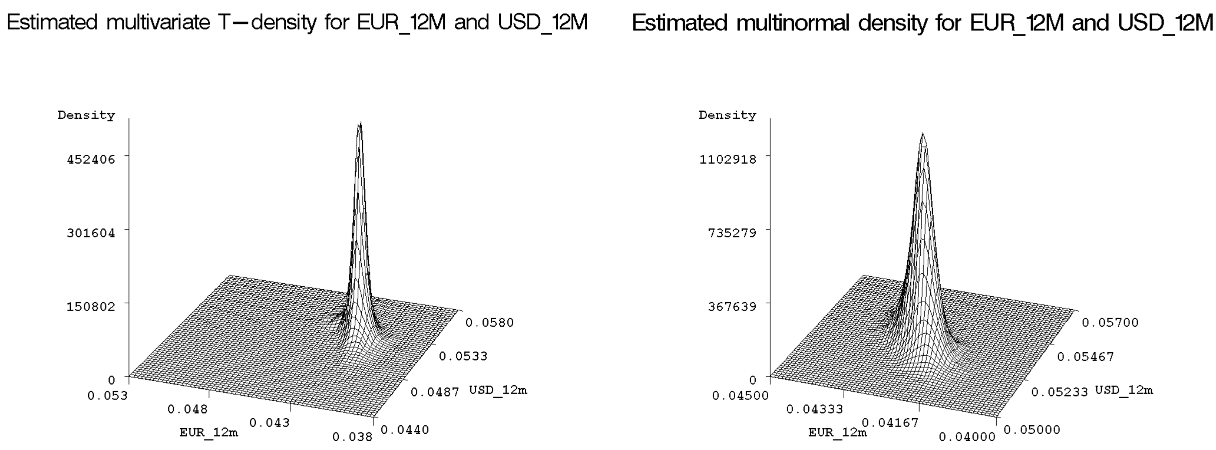


Figure 5: Joint PDF of the EUR 12M and USD 12M (error vectors have multiv.t- and normal-distribution)