# A MECHANICAL APPROACH OF MULTIVARIATE DENSITY FUNCTION APPROXIMATION

**László Mohácsi[(a)], Orsolya Rétallér[(b)]**

[(a)] Department of Computer Science, Corvinus University of Budapest
[(b)]MTA-BCE "Lendület" Strategic Interactions Research Group and
Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest

[(a)] mohacsill@gmail.com, [(b)]retaller.orsolya@gmail.com

## ABSTRACT
In this paper we are modeling multivariate density functions by going back to the roots: instead of trying to fit a well-known copula on the data, we choose to generate one. Our model approximates the two dimensional density function using physical analogy. Points on the scatter plot diagram are represented by small balls of unit mass placed on a sheet of elastic sponge, and the deformation of the surface of the sponge caused by the balls represents the density of the points on the scatter plot diagram. The elasticity of the sponge is described by Hooke's law. The distortion of the sponge can be determined by using finite element methods. The distorted surface can be approximated by functions using Fourier transformation. Hence, the model can be extended into higher dimensions.

Keywords: copulas, multivariate data analysis, finite element method, goodness of fit

## 1. INTRODUCTION
Modeling multivariate distributions is relatively complicated in general, even when we are aware of the marginal distributions' nature. The dependency structure among the variables is usually described by their covariance or correlation matrix, however these measures have a great disadvantage: they only measure the linear dependency, not the association in general. There are other measures, such as Spearman's rho, or Kendall's tau, which do not rely heavily on linearity, but these measures are rather used to determine how monotonic the relation is among the given variables. But we should note, that dependency doesn't imply monotonicity.

Copulas are functions that join multivariate distribution functions to their one-dimensional margins. (Nelsen 2006 and Sklar 1959) They also take it into consideration that the dependency might be different on the edges, they provide a flexible structure, and do not intend to measure linearity or monotonicity.

There are several famous copulas, which we might try to fit on the data. The problem is, that in some cases none of the famous copulas fit well – even if we take the parameter shifting into consideration. This was basically what motivated us trying to find alternatives for these situations.

In section 2 we introduce our copula generating method, for which we are using finite element methods. This is a modeling technique widely used among engineers, but as far as we're concerned it has no previous history in financial modeling. We believe, that the synergy of the different fields will provide with an interesting and promising result. In section 3 we present some calculations on financial data in two cases. Not only we would like to show how our copula fits on the data as one of the bests (compared to some of the famous copulas), but we would also like to demonstrate, that the conventional chi-square testing of the goodness-of-fit is not reliable in our case. In section 4 conclusion follows.

## 2. THE COPULA GENERATING METHOD
In order to understand how our copula generation works let's consider an analogy from physics. The density function of a given copula can be modeled by the distortion of an elastic lattice (sponge) when small balls of unit mass are placed on it. The more balls we place on a given point, the more the sponge sinks. Each site , where we decide not to put any balls the value of this function is zero, or – due to the interaction of the balls – close to zero. If we would like to obtain a function that has the characteristics of a density function (e.g. non-negativity) this procedure has to be reversed: instead of placing weights on the sponge, and pushing it down, we prefer to pull it up. Hence, the surface looks more like the surface of a density function, and fulfills the non-negativity criteria.

The distortion of the lattice can be determined by using finite element methods. In engineering, these methods are typical modeling applications of the distortion of statically loaded machine parts. Another typical application is determining the deformation of machine parts and temperature distribution in them with a given thermal conductivity and thermal expansion coefficient. In this application, temperature dependency is omitted , however it could lead to an extension of modeling in 3D.

For generating the 3D model, we used our own program written in C#. The Z88 Aurora is a program, which is a user interface based on the Z88 engine (http://www.z88.de/).

### 2.1. The parameters of the simulation

During the simulation the behavior of the sponge can be characterized with the following parameters:

1. The *Young-modulus* (E, N/mm$^2$), also known as the elastic modulus , which is a constant in our model. It describes the relationship between the force effecting on the lattice and its' distortion, which is also known as Hooke's law. However, it should be noted that not all materials are behaving according to Hooke's law's : e.g. amorphous materials rubber the distortion and the force on it does not have a linear dependency.
2. The *Poisson's ratio* is a dimensionless constant. It provides the negative ratio of transverse to axial strain.
3. The next parameter of the material is the *density* (ρ, kg/m$^3$). If we place the material in the gravitational field, it will be distorted by its own weight.
4. *Thermal conductivity* (λ, W/(m·K)) could also be taken into consideration. According to the second law of thermodynamics, an isolated system, if not already in its state of thermodynamic equilibrium, spontaneously evolves towards it. For non-evolving materials, this phenomenon is characterized by the Fourier-law.
5. There are a few materials which have a *negative coefficient of thermal expansion*, which means that cooling them leads to expansion. An example for this phenomenon is water between 0 C° and 4 C°.

### 2.2. The simulation steps

The simulation comprises the following steps. (Based on Deák 1990 and Ross 1997.)

#### 2.2.1. Building the model

As a first step we have to build the model of the shape that we would like to observe. The basis of the model relies on the nodes, which are given by their coordinates in 3D. These nodes determine spatial elements, which are usually constructed of hexahedrons or tetrahedrons. Eight nodes could determine a hexahedron, while six of them lead to a tetrahedron.

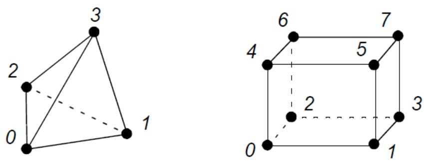


Figure 1: A tetrahedron and a hexahedron given by their nodes.

In engineering there is a common practice regarding the model building: on those areas, where the forces are more concentrated there should be more dense sampling.

Most finite element softwares are capable of decomposing models, which originate from CAD programs. In our case it was more desirable to generate the nodes and elements from the program.

#### 2.2.2. Boundary conditions

Before running the simulation we have to provide the boundary conditions. At least one node should be designated, which has a fixed position. Without this step a static examination is unconceivable. Also, for each node the attacking force vectors can be given.

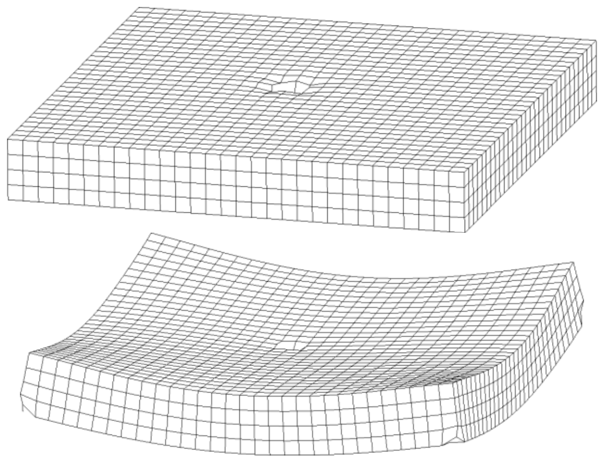

Figure 2: A loaded rubber sheet, which is supported in every point and in the middle only.

#### 2.2.3. Running the simulation

As we expected the value of the Young modulus had no significant effect on our simulations, and it was also irrelevant, whether we pushed or pulled the material. We found however, that the thickness of the material is relevant. Also, we had an interesting side-effect, which we decided to call the sponge-effect.

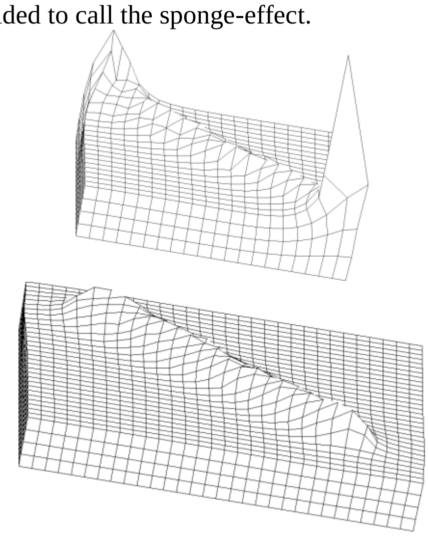

Figure 3: The sponge-effect.

The sponge-effect can be explained again with an analogy deriving from physics. Consider a mattress, and imagine if we sit on it. If we sit in the middle, there will be a big distortion in the middle, some distortion in the surrounding area, but almost no distortion on the edges. However, if we decided to sit on the edge, the distortion would be much bigger, as the surrounding area is missing, and there is no support from it. Therefore, if we use the same force to pull the sponge in the middle and on the edges, the distortion looks completely different.

As the sponge-effect distorted our simulations a lot, we decided to obtain some corrections. Figure 3 shows that we decided to put some extra elements outside our model as well, which gives some support to the nodes on the edges as well.

## 3. EXAMPLES AND COMPARISON

We applied the previously described methodology on two datasets, both of them are modeling dependency among certain financial indicators. In the first case we observed the stock exchange indices of London's and Paris's stock exchange market, namely the FTSE and FCHI. The second case is about the dependency between two financial assets: gold and real estate.

### 3.1. The FTSE and FCHI dependency

The data we used to observe the dependency between the English and the French market were the daily closing values of the indices between 01.01.2000 and 12.31.2009, but we have only taken into consideration those days, where both markets were open. As a result we obtained 2412 pairs of data.

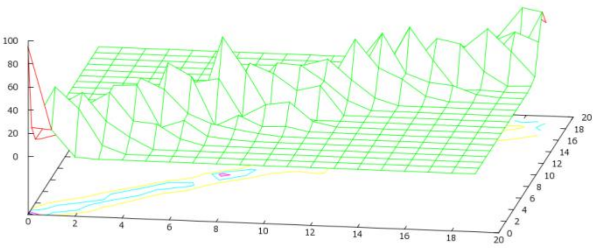

Figure 4: The relationship between ranking numbers on a 20x20 crosstabulation

As we originally expected, the FTSE and FCHI indices have a very high positive correlation, namely 0.968. Figure 4 represents the relationship of the ranking numbers on a joint histogram on a 20x20 crosstab. There is not much difference if we use the original data or the ranking number. This also means that the Pearson's correlation and Spearman's rho both provide us with similar results and these measurements provide enough information about the dependency structure. Still, let's observe which is the best fitting copula on the data.

Both in this case and the next one we fitted the independence copula along with Clayton's and Gumbel's copulas (the formulas can be found in table 1), and of course our one. For those copulas that require a parameter estimation (Clayton's and Gumbel's) we applied the maximum likelihood estimation using Excel Solver. As a result we ended up with a theta value of 7.540 for Clayton's copula, and 4.331 for Gumbel's.

Table 1: CDF of the fitted copulas

| Copula name | Bivariate formula of the CDF |
|---|---|
| Independence | $uv$ |
| Clayton's | $\left(max\{u^{-\theta}+v^{-\theta}-1;0\}\right)^{-1/\theta}$ |
| Gumbel's | $exp\left(-\left[\{-log(u)\}^{\theta}+\{-log(v)\}^{\theta}\right]^{\frac{1}{\theta}}\right)$ |

It should be noted, that instead of providing a figure of the cumulative distribution function of the copulas we decided to represent the expected values. This way it is easier to compare the results with the original data.

To calculate the expected value of each cell we used the following formula (i denotes the number of row, and j stands for the number of column, k is dimension of the crosstab, n denotes the number of data pairs, and F stands for the cumulative distribution function of the given copula):

$$n\cdot\left(F\left(\frac{i}{k},\frac{j}{k}\right)-F\left(\frac{i-1}{k},\frac{j}{k}\right)-F\left(\frac{i}{k},\frac{j-1}{k}\right)+F\left(\frac{i}{k},\frac{j}{k}\right)\right)$$

Figure 5-8 show the results of the fitting.

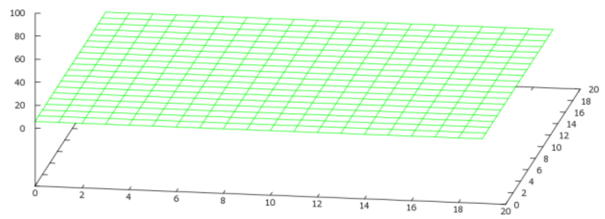

Figure 5: The expected values based on the independence copula

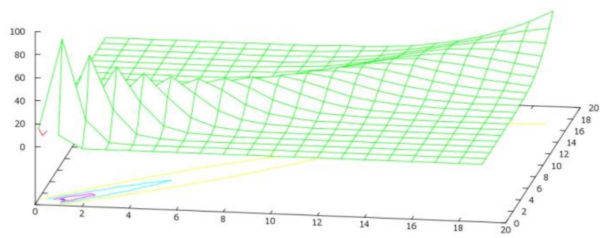

Figure 6: The expected values based on Clayton's copula (θ = 7.540)

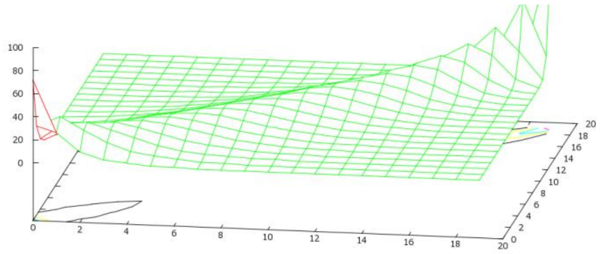

Figure 7: The expected values based on Gumbel's copula (θ = 4.331)

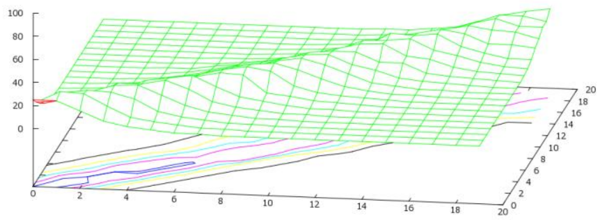

Figure 8: The expected values based on our copula

It is not easy to decide by looks which is the best fitting copula on these data. However it is quite visible, that the independence copula's case seems to be the worst one.

With the statistical testing we encountered a major difficulty, which was the fact, that the conventional chi-square method requires an expected value of at least 5 in each cell to be able to perform the test. In general if this criterion isn't met, the suggestion is to merge some of the cells, until the test can be calculated. In this case however – as the data are basically in the diagonal – even the 3x3 representation is inadequate for the statistical testing.

The goodness-of-fit testing of copulas is still a hot topic, and there is not yet an obvious answer how it should be done. (As an example see Berg 2007, Dobric and Schmid 2005, Lowin 2001, Patton 2006, Quessy 2005). This paper does not have an agenda to provide an answer to this question hence we only calculated the chi-square test (if it was possible) and the average sum of squares of the errors. In the previous case it should also be noted, that we disregarded the test's requirement regarding the minimal expected value (because of the previously described phenomenon), but this caused major distortion in the test value in those cases where the expected value was close to zero. If it was even zero, the test couldn't even be calculated. This was basically the reason why we decided to observe the average sum of square of the errors instead. Table 2-7 represent the results, and the best ones are highlighted.

It should also be noted, that the results vary if we change the numbers of cells in the crosstabs. In this case we did the calculations on a 3x3, a 5x5 and a 20x20 crosstab.

Table 2: Chi-square results (3x3 crosstab)

| Fitted copula | Chi-square | df | p-value |
|---|---|---|---|
| Independence | 3 720,0 | 8 | 0,000 |
| Clayton's | 140,2 | 7 | 0,000 |
| Gumbel's | 239,6 | 7 | 0,000 |
| Our one | 2 078,7 | 7 | 0,000 |

Table 3: Chi-square results (5x5 crosstab)

| Fitted copula | Chi-square | df | p-value |
|---|---|---|---|
| Independence | 5 957,5 | 24 | 0,000 |
| Clayton's | 348,3 | 23 | 0,000 |
| Gumbel's | 384,0 | 23 | 0,000 |
| Our one | N/A | N/A | N/A |

Table 4: Chi-square results (20x20 crosstab)

| Fitted copula | Chi-square | df | p-value |
|---|---|---|---|
| Independence | 11 923,3 | 399 | 0,000 |
| Clayton's | 1 616,0 | 398 | 0,000 |
| Gumbel's | 2 939,0 | 398 | 0,000 |
| Our one | N/A | N/A | N/A |

Table 5: Average Sum of Squares (3x3 crosstab)

| Fitted copula | Average Sum of Squares |
|---|---|
| Independence | 110 773,1 |
| Clayton's | 3 462,6 |
| Gumbel's | 6 259,7 |
| Our one | 65 873,2 |

Table 6: Average Sum of Squares (5x5 crosstab)

| Fitted copula | Average Sum of Squares |
|---|---|
| Independence | 22 991,0 |
| Clayton's | 2 230,3 |
| Gumbel's | 2 319,8 |
| Our one | 2 922,5 |

Table 7: Average Sum of Squares (20x20 crosstab)

| Fitted copula | Average Sum of Squares |
|---|---|
| Independence | 179,7 |
| Clayton's | 52,2 |
| Gumbel's | 84,5 |
| Our one | 75,3 |

This case was a difficult one, as none of the fitted copulas were good enough to fit according to the chi-square test results. However, the figures we presented can be quite persuasive, that the expected values generated based on Clayton's, Gumbel's and our copula, are all pretty close to the original data.

### 3.2. The gold and real estate dependency

In this case 2043 pairs of data have been observed for a 10 year period again. This case the dependency was not that obvious, the Pearson's correlation coefficient was only -0.442. But if we look at the data we can find certain groups on the scatter-dot diagrams. Figure 9 represents the original data and the rankings as well, whereas figure 10 presents the joint histogram on a 20x20 crosstabulation.

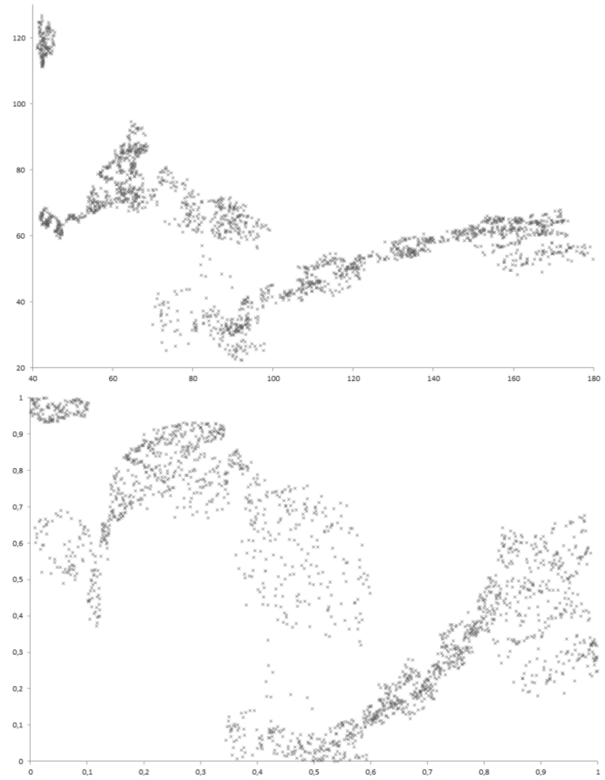

Figure 9: The relationship between the original data and the ranking numbers

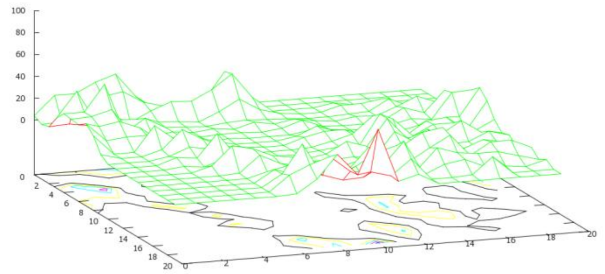

Figure 10: The relationship between ranking numbers on a 20x20 crosstabulation

As the maximum likelihood estimation didn't bring any results for neither Clayton's nor Gumbel's theta, we decided to use -0.2 for the first, and 1 for the second one (these estimations had pretty good results for the sum of the log-likelihoods). This resulted, that the Gumbel copula was identical to the independence one. The fitted copulas can be seen on figure 11-14.

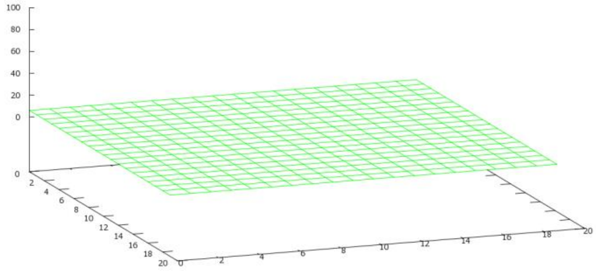

Figure 11: The expected values based on the independence copula

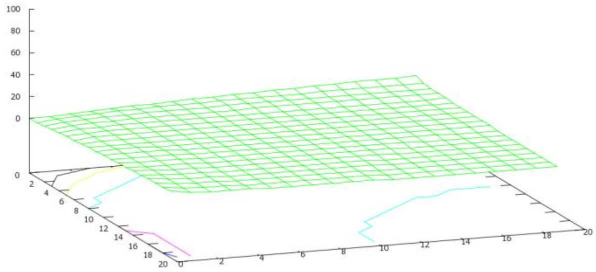

Figure 12: The expected values based on Clayton's copula (θ = -0.2)

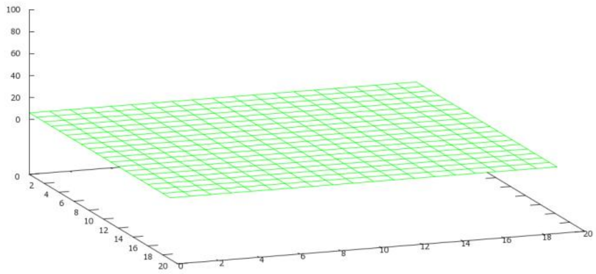

Figure 13: The expected values based on Gumbel's copula (θ = 1)

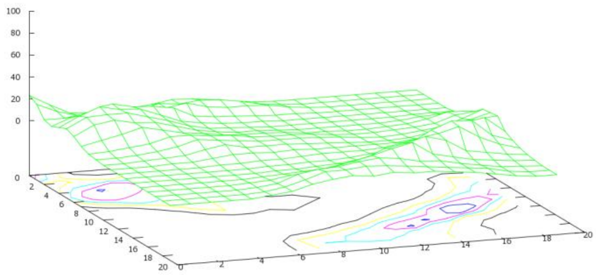

Figure 14: The expected values based on our copula

It is quite visible, that in this case our copula is the closest to the original data. Once again, we tried to obtain the conventional chi-square test to verify this statement, but we found that if we want to follow the rules, a 3x3 crosstab is the biggest that we can use. In this case however the test could really be applied, as the expected value in each cell was over 5 in all cases. Still, none of the fitted copulas were good enough to accept the null hypothesis. Table 8-13 represent the results.

Table 8: Chi-square results (3x3 crosstab)

| Fitted copula | Chi-square | df | p-value |
|---|---|---|---|
| Independence | 796,8 | 8 | 0,000 |
| Clayton's | 1 291,7 | 7 | 0,000 |
| Gumbel's | 796,8 | 7 | 0,000 |
| Our one | 1 344,3 | 7 | 0,000 |

Table 9: Chi-square results (5x5 crosstab)

| Fitted copula | Chi-square | df | p-value |
|---|---|---|---|
| Independence | 1 605,3 | 24 | 0,000 |
| Clayton's | 2 278,3 | 23 | 0,000 |
| Gumbel's | 1 605,3 | 23 | 0,000 |
| Our one | 1 652,6 | N/A | N/A |

Table 10: Chi-square results (20x20 crosstab)

| Fitted copula | Chi-square | df | p-value |
|---|---|---|---|
| Independence | 5 470,1 | 399 | 0,000 |
| Clayton's | 8 717,2 | 398 | 0,000 |
| Gumbel's | 5 470,1 | 398 | 0,000 |
| Our one | N/A | N/A | N/A |

Table 11: Average Sum of Squares (3x3 crosstab)

| Fitted copula | Average Sum of Squares |
|---|---|
| Independence | 23 725,8 |
| Clayton's | 35 301,4 |
| Gumbel's | 23 725,8 |
| Our one | 4 969,3 |

Table 12: Average Sum of Squares (5x5 crosstab)

| Fitted copula | Average Sum of Squares |
|---|---|
| Independence | 6 195,0 |
| Clayton's | 8 039,9 |
| Gumbel's | 6 195,0 |
| Our one | 3 443,8 |

Table 13: Average Sum of Squares (20x20 crosstab)

| Fitted copula | Average Sum of Squares |
|---|---|
| Independence | 82,5 |
| Clayton's | 91,2 |
| Gumbel's | 82,5 |
| Our one | 80,0 |

As the chi-square testing is very unreliable, we prefer to make a decision on the average sum of squares of errors. In this case our copula seems to provide with the best solution. However, the goodness-of-fit testing for copulas is still an interesting topic with a lot of open questions.

## 4. CONCLUSION

Even though the results are not conclusive (because of the unreliability of the testing), we believe that the copula generation method we presented could be widely applied. It is flexible enough to fit even on those data, where other copulas cannot find any relationship among the data. Also, as we have referred to it, it could be extended into higher dimensions.


## ACKNOWLEDGMENTS

The authors would like to thank Erzsébet Kovács for encouraging them to work together, even though they have very different fields of expertise. They would also like to acknowledge Zsolt Benczúr for his helpful hints, and Fruzsina Gajdos for her help with the translation of the professional expressions.


## REFERENCES


Berg, D., 2007. Copula goodness-of-fit testing: an overview and power comparison. Available from: https://www.duo.uio.no/bitstream/handle/10852/10361/stat-res-05-07.pdf?sequence=1

Deák, I., 1990. *Random number generators and simulation (Mathematical Methods of Operations Research)*. Budapest: Akadémiai Kiadó.

Dobric, J., Schmid, F., 2005. Testing Goodness of Fit for Parametric Families of Copulas – Application to Financial Data. *Communications in Statistics – Simulation and Computation*, 34: 1053-1068.

Klubman S.A., Panjer, H.H., Willmot G.E., 2008. *Loss Models: From Data to Decisions*. Hooken, New Jersey: John Wiley & Sons.

Lowin, J.L., 2001. The Fourier Copula: Theory & Applications. Harvard University.

Nelsen, R.B., 2006. *An Introduction to Copulas*. New York: Springer

Panjer, H.H., 2007. *Operational Risk: Modelling Analytics*. Hooken, New Jersey: John Wiley & Sons.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47 (2): 527-556.

Quessy, J.F., 2005. *Méthodologie et Application des Copules: Tests d'Adéquation, Tests d'Indépendance, et Bornes pour la Valeur-à-Risque*. Thesis (PhD). Université Laval.

Ross, S.M., 1997. *Simulation*. Boston: Academic Press.

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