Modeling Risk Factor Interaction Using Copula Functions

Peter Bidyuk[1[0000-0002-7421-3565]], Olexander Trofymchuk [2[0000-0003-3782-4209]], Irina Kalinina[3[0000-0001-8359-2045]], and Aleksandr Gozhyj [3[0000-0002-3517-580X]]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,
[2] Institute for Telecommunications and Global Information Sphere at the National Academy of Sciences of Ukraine, Kyiv, Ukraine,
[3] Petro Mohyla Black Sea National University, Nikolaev, Ukraine

pbidyuke_00@ukr.net, itelua@kv.ukrtel.net, irina.kalinina1612@gmail.com,alex.gozhyj@gmail.com

Abstract. The procedure is proposed for analyzing the risk factor interaction in financial systems. The procedure is based upon the results of eigenvalues distribution analysis and distances between the eigenvalues for empirical and theoretical dependency matrices. Some results of the theory of random matrices are used to interpret the results achieved in the process of empirical studies for the correlation matrices of different kind. The results of computational experiments show that for small eigenvalues the results of theoretical analysis for random matrices are similar to the empirical matrices. The number of eigenvalues that exceed theoretical thresholds corresponds to the principal factors in a model. The difference between theoretical and empirical distributions of distances between eigenvalues means that in practice there almost always exist a large eigenvalue indicating (in economic interpretation) on existence of dominating generalized market factor. It was also established that no extra internal influence factors exist when the widely used models of derivative costs are hired. This result provides a possibility for determining correctly the number of principal factors to construct mathematical models necessary for practical applications.

Keywords: risk factor, financial systems, correlation coefficients, correlation matrices, multivariate model

1 Introduction

Modern financial instruments are basically characterized as nonlinear non-stationary processes functioning in complex conditions of multiple stochastic disturbances. Such conditions require development and application of non-traditional mathematical models for adequate describing the processes necessary for solving the tasks of forecasting, risk estimation and decision making. This is especially true to financial risk analysis in a case of multivariate problem statement. Simultaneous influence of risk factors and their interaction may result in much higher losses than available simplified models indicate that do not take into account possible interactions. That is why modeling of risks in the problems of risk management especially in a case of large scale systems should not be limited to analysis of separate factors. Such models should also take into consideration possible interactions between the risk factors [1-3]. The cost of some financial instrument (position) usually depends on a set of exogenous risk factors that are generated by the current state of an enterprise, economy branch or macroeconomy as a whole. On the other side they are endogenous regarding to the functioning of specific market. Emerging of the endogenous factors is a feature of self-organization in large scale systems [4]. Larger number of risk factors results in higher frequency of extreme events and in distributions with heavier tails. At the same time existence of links between the cost of financial instrument and the relevant risk in practice does not produce noticeable tendency to income growth with growing risks.

The models constructed by selecting some (usually not large) number of principal factors of influence are popular in economy and finances thanks to simplicity of constructing procedure and their convenient low dimension. When modeling specific markets usually the following external factors are selected: general market factor, specific economy branch factors, and the factors that influence directly the market position. For example, relationship between nominal and market stock price, as well as characteristics of specific transaction such as its volume, conditions of payment etc.

The group risk model for stock markets was proposed in the study [5]. According to the assumption of the model the market consists of several separate groups that include specific financial instruments the prices of which are correlated with the other stock prices belonging to the same group. Usually when model constructing for complex systems is performed the principal risk factors are selected on the basis of existing economic theory. An alternative approach is based on detecting of available risk factors using mathematical techniques [6]. In this study we propose to characterize dependences between the stock prices using the results of random matrix theory [7] by application of the results to the matrices of dependency measures and coordination between the systems of financial instruments. The study [8] is based on analysis of concentration of especially large eigenvalues of random symmetric matrices. In the study [9] it was found the correspondence between eigenvalues distribution with the theoretical results from the theory of random matrices using empirical matrices of linear correlations for 406 stock prices of selected USA companies in the period of time 1991-1996 without taking into consideration the 6% of the largest eigenvalues.

The study [10] stresses the correspondence between the results achieved for the symmetric random matrix and distribution of distances between the eigenvalues of empirical matrix of linear correlations for 1000 US stocks within two year period. In the work [11] the method is proposed for cleaning the noise from the empirical matrix of linear correlation coefficients.

The linear correlation coefficients exhibit some drawbacks regarding application in risk management systems. That is why we considered application of some results from the random matrices theory to analysis of risk measures dependency. Taking into consideration the necessity of determining effectiveness of application the methods of random matrices analysis to risk management problems the multivariable model of financial instruments was constructed with known dependency structure. Further on we studied how well the results achieved for non-random empirical matrices with adding some noise are similar to the theoretical results relevant to random matrices. An important problem related to the factor model building is establishing the links between numerical description of the dependency and the number of principal factors.

2 Statement of problem

The purpose of the study includes solving the following tasks: (1) constructing multivariate statistical model with known mutually dependent principal factors on the basis of copulas and distribution functions; (2) studying the possibilities for application of the methods related to random matrices theory to analysis of the dependency measures; (3) analyzing the possibility of application the eigenvalue distribution of correlation matrices related to different dependency measures and distribution of distances between the eigenvalues with the final purpose of determining the number of principal factors in a model.

3 Estimation of risk dependencies

The proper probabilistic description of a system including a set of stochastic processes is the following probability: $P(X_1 \leq x_1; \ldots; X_n \leq x_n)$, i.e. their joint probability distribution, $(x_1, \ldots, x_n)$. The distribution contains information related to the processes dependency structure and marginal distributions of each random variable. To distinguish between descriptions of dependency between these random variables the special link functions can be used known as copulas.

Definition 1: the function, $C: [0,1]^n \to [0,1]$, is called *n*-copula if the following conditions are hold:

1. $C(F_1, \ldots, F_n) = 0$, if there exists j such that $F_j = 0$;
2. $C(1, \ldots, 1, F_i, 1, \ldots, 1) = F_i$;
3. C is *n*-increasing function.

Theorem 1 [12]: Let H is *n*-dimensional joint distribution function with marginal distributions, $\mathrm{F}_1, \ldots, \mathrm{F}_\mathrm{n}$. Then there exists such *n*-copula *C* that for all $\vec{\mathrm{x}} \in \mathrm{R}^\mathrm{n}$ the following equality holds:

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

If the functions, $\mathrm{F}_1, \ldots, \mathrm{F}_\mathrm{n}$, are continuous then, *C*, is unique; otherwise, the functions, *C*, are uniquely determined on the $\mathrm{Rng}[\mathrm{F}_1] \times \mathrm{Rng}[\mathrm{F}_\mathrm{n}]$. And vice versa: when, $\mathrm{F}_1, \ldots, \mathrm{F}_\mathrm{n}$, are distributions and *C* is *n*-copula, then, $\mathrm{H}(\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n})$ is *n*-dimensional joint distribution function with marginal distributions, $\mathrm{F}_1, \ldots, \mathrm{F}_\mathrm{n}$.

Thus, copula is sufficiently (completely) defining the dependency structure between the random variables selected. However, the direct practical application of copulas to description of tens or hundreds random variables meets definite difficulties. First, selection of appropriate copula family suitable for determining the copula forming the dependency structure by some parameter estimation is rather difficult. When the dimensionality is high the number of copula parameters is growing substantially and the problem of parameter estimation arises due to incomplete observations, for example, maximum likelihood procedure for parameter estimation may not work. If risk measures are estimated for high dimensional copulas the Monte Carlo procedures cannot be effective.

To solve the high dimension dependency modeling problem it is proposed to detect principal factors in the manner as it is being done in the economic and financial model building procedures. Then the influence factors can be modeled via copulas and marginal distributions, and appropriate dependency measures. Though here the problem comes to being of determining the number of principal variables for which the joint distribution is constructed.

There are several specific features that are desirable for a dependency measure to possess [13]:

1. it should be defined for any pair of continuous or discrete random variables X and Y;
2. the measure should be symmetric;
3. it should be equal to zero in a case of independent random variables;
4. it should be limited to the range of [–1; 1] and reach the lowest and the highest values when both random variables are, respectively, counter monotonic and equally monotonic;
5. it can be expressed via the Pearson linear correlation coefficient in a case of two-dimensional normal distribution;
6. it distinguishes not only between random variables but also provides a measure of distance between them;
7. it should be invariant relatively continuous strictly increasing transforms.

If a measure exhibits all the features mentioned above it is called the dependence metrics (measure).

One of the widely used dependency measures when modeling multivariate risks with elliptical distributions is linear correlation. For example, it is used in the case of normal and t-distributions. The linear correlation coefficient between two random variables X and Y with finite standard deviations is determined via the expression:

$$\rho(X,Y) = \frac{E[XY]-E[X]E[Y]}{\sqrt{\sigma^2[X]\sigma^2[Y]}}, \tag{2}$$

where, $\sigma[X]$ and $\sigma[Y]$, are standard deviations for X and Y, respectively.

The linear correlation coefficient cannot be the dependency metric because there exist distributions with infinite standard deviation, and expression (2) is not defined in such cases. It means that the first characteristic defined above does not exist for the coefficient. The coefficient of linear correlation is commutative, i.e. $\rho(X,Y) =$

$\rho(Y, X)$. In the case of independent random variables the following condition holds: $\rho(X, Y) = 0$, though it does not follow from the equality, $\rho(X, Y) = 0$, that the variables are independent. The coefficient of linear correlation is limited by the values: $-1 \leq \rho(X, Y) \leq 1$, and equality is reached when the two random variables are completely dependent. The coefficient is invariant to strictly increasing linear transforms, though in general case it is not invariant to nonlinear strictly increasing transforms.
The random variables are considered to be in concordance when a tendency exists to simultaneous increasing or decreasing of their values. The observations, (x_i, y_i), and, (x_j, y_j), belonging to the vector or random variable, (X,Y), are considered to be in concordance if $x_i < x_j, \ \ y_i < y_j$ or $x_i > x_j, \ \ y_i > y_j$. The observations concordance condition can also be written in the way: $(x_i - x_j)(y_i - y_j) > 0$; and the non-concordance condition is as follows: $(x_i - x_j)(y_i - y_j) < 0$.

The Kendall rank correlation, τ, and the rank Spearman correlation coefficient, ρ_S, are numerical characteristics of dependency that are related to the concordance measures. The Kendall, τ, is concordance measure for a sample of two random variables, X and Y, which is calculated as a difference between the number of coordinated and non-coordinated pairs of two-dimensional observations divided by the general number of pairs of the two-dimensional observations. Let, (X', Y') and, (X'', Y''), are independent random vectors having the same joint distribution functions. Then, for the general sample of the random vector components with such joint distribution the concordance measure of Kendall τ is written as follows:

$$\tau = P[(X' - X'')(Y' - Y'') > 0] - P[(X' - X'')(Y' - Y'') < 0].$$

For the increasing transforms ψ, ϕ with $X' \geq X''$ the following inequality holds: $\psi(X') \geq \psi(X'')$, and for $Y' \geq Y''$ the following condition holds: $\phi(Y') \geq \phi(Y'')$. Thus, according to definition, we have invariance of Kendall τ to increasing transforms.

For the two-dimensional normal distribution as well as for any other random variable exhibiting dependency structure described by elliptical copula, the Kendall τ can be expressed via linear correlation coefficient, ρ, as follows:

$$\tau = \frac{2}{\pi} \arcsin \rho.$$

The Spearman coefficient of rank correlation, ρ_S, is also related to the notions of concordance and non-concordance. But the measure also takes into consideration marginal distributions of random variables. Let, (X', Y'), (X'', Y''), and, (X''', Y''') are independent random vectors having the same joint distribution functions, then the Spearman coordination measure, ρ_S, is defined as follows:

$$\rho_S = P[(X' - X'')(Y' - Y''') > 0] - P[(X' - X'')(Y' - Y''') < 0].$$

The dependency measure can be defined in the same way via another component of some third vector, X'''.

The Spearman rank correlation can be expressed via linear Pearson correlation coefficient, ρ, as follows [12]:

$$\rho_S = \frac{E(FG)-1/4}{1/12} = \frac{E[FG]-E[F]E[G]}{\sqrt{\sigma^2[F]\sigma^2[G]}} = \rho(F,G),$$

where, *F* and *G* are marginal distribution functions for the random variables *X* and *Y*, respectively.

The coefficients of rank correlation, τ, and, ρ_S, are commutative: $\tau(X,Y) = \tau(Y,X)$, $\rho_S(X,Y) = \rho_S(y,X)$. For completely independent random variables, $\tau(X,Y) = \rho_S(X,Y) = 0$. The values of both rank correlation coefficients belong to the interval: [-1, 1]. These concordance measures can be expressed via copulas.

Theorem 2 [14]: If *X* and *Y* are continuous random variables with joint distribution function *H* and marginal distribution functions *F* and *G*, respectively, and *C* is a copula such that $H(x,y) = C\big(F(x),G(y)\big)$, then Kendall rank correlation coefficient is defined as follows:

$$\tau = 4\int_{[0,1]^2} C(u,v)dC(u,v) - 1,$$

and Spearman rank correlation coefficient can be computed via the expression:

$$\rho_S = 12\int_{[0,1]^2} u,vdC(u,v) - 3 = 12\int_{[0,1]^2} C(u,v)dudv - 3.$$

Thus, if for the two pairs of random variables (X_1,Y_1) and (X_2,Y_2) the dependences between which have a copula form of, C_1, and C_2, respectively, and such that the following inequality holds:

$$C_1(u_1,u_2) \geq C_2(u_1,u_2),\ \forall u_1,u_2 \in [0;1],$$

then, the concordance measure for the pair of random variables (X_1,Y_1) is greater than for the pair of variables (X_2,Y_2).

4 The matrices of correlation coefficients

A dependence measure characterizes the dependence structure between two random variables with one number. Generalization of the measure for the case of *N* >2 random variables is *N*×*N* matrix of paired dependency measures. The correlation matrix can be theoretical and empirical that is used in practice. For example, empirical linear correlation matrix is a key part of the model for estimation of the Value-at-Risk measure for the normally distributed risks, and Markowitz optimal portfolio corresponds to small eigenvalues of the correlation matrix [15, 16].

For the model of mean correlations when all elements of the correlation matrix are equal to, ρ, but for the "1-s" on the main diagonal, there is one large eigenvalue, $\lambda_1 = 1 + (N-1)\rho$, and all other eigenvalues are equal to $\lambda_{i\geq 1} = 1 - \rho$. A similar result was achieved in the case when all non-diagonal elements of correlation matrix are random values with expectation, ρ, and standard deviation, σ:

$$E[\lambda_1] = (N-1)\rho + \frac{\sigma^2}{\rho} + 1 + o(1).$$

Thus, when, $\rho > 0$, the largest eigenvalue is increasing with growth of a system dimensionality N. The dominating eigenvalue corresponds to equally distributed over its component's eigenvector, $\nu_1 = (1/\sqrt{N})(1,1,\ldots,1)$. This vector has an economic importance as a factor that influences simultaneously all financial positions or generalized market index. The factor can be hired to explain, for example, high generalized market crises. Such interpretation finds a support in the studies of empirical financial correlation matrices [10, 17]. However, in practice we observe availability of several eigenvalues in the interval the order of which is overcoming by 5 – 10 times the basic part of the matrix eigenvalues. This fact can be explained by influence of not only generalized market factor but by the separate branch factors too that influence on some part of the positions available. In such cases the correlation matrix approaches to the block-diagonal one each block of which corresponds to the specific branch of economy. Usually the correlations within the block are higher than the correlations outside of the block.

Such situation can be characterized by the matrix containing $N_1 \times N_1$ blocks on the main diagonal with the following values of correlations: "1-s" on the diagonal; ρ_1 for the non-diagonal elements, and ρ_0 outside of the blocks. The highest eigenvalue of the correlation matrix is determined in such case by the expression:

$$\lambda_1 = 1 + (N_1 - 1)\rho_1 + (N - N_1)\rho_0.$$

There are eigenvalues that correspond to the eigenvectors that characterize basic influence of the economy branch:

$$\lambda_{i=2\ldots\frac{N}{N_1}} = 1 + (N_1 - 1)\rho_1 - N_1\rho_0;$$

and other eigenvalues:

$$\lambda_{i=\frac{N}{N_1}+1\ldots N} = 1 - \rho_1.$$

To determine statistical characteristics of the correlation matrix eigenvalue spectrum the results of the random matrices theory can be used. The theory was developed in 1950-s for the needs of physicists who studied the complex quantum system spectra. The matrices of Pearson, Kendall, and Spearman correlation coefficients are symmetric what is suitable for considering the case of maximum statistical independence that can be reached in symmetry conditions. The possible deviations from the random matrices theory point out to existence of specific dependences for the systems under consideration.

Theorem 3 [7]: Let H is real-valued symmetric random $N \times N$ matrix with non-diagonal elements, $H_{i,j}, i > j$, that are zero mean independent and identically distributed with nonzero standard deviations. Then distribution density of the random matrix H is defined as follows:

$$P(H) = \exp(-a \cdot trH^2 + b \cdot trH + c), \quad (3)$$

where, a>0, b and c are real constants. The N×N symmetric random matrix is completely characterized by $N(N+1)/2$ random values that determine all, ρ_{ij}. Recollect that the eigenvalues, $\lambda_1, \dots, \lambda_N$, of a random matrix are also random. Taking into consideration that the elements of the right-hand side of expression (3) are determined as

$$tr\, H^2 = \sum_1^N \lambda_j^2, \qquad tr\, H = \sum_1^N \lambda_j,$$

we have:

$$P\left(\lambda_1, \dots, \lambda_N; \nu_1, \dots, \nu_{\frac{N(N-1)}{2}-N}\right) = exp\left(-a \cdot \sum_1^N \lambda_j^2 + b \cdot \sum_1^N \lambda_j + c\right) J(\vec{\lambda}, \vec{\nu}), \quad (4)$$

where, $\nu_1, \dots, \nu_{\frac{N(N-1)}{2}-N}$ are independent random values, that together with the eigenvalues, $\lambda_1, \dots, \lambda_N$, define random matrix, Jacobean:

$$J(\vec{\lambda}, \vec{\nu}) = \left| \frac{\partial(H_{11}, \dots, H_{NN})}{\partial(\lambda_1, \dots, \lambda_N; \nu_1, \dots, \nu_{\frac{N(N-1)}{2}-N})} \right|.$$

It is shown in [7] that for symmetric random matrix such Jacobean can represented in the form of a product of eigenvalues function and the function of introduced parameters, $\nu_1, \dots, \nu_{\frac{N(N-1)}{2}-N}$, as follows:

$$J(\vec{\lambda}, \vec{\nu}) = f(\vec{\nu}) \prod_{1 \le j < k \le N} |\lambda_k - \lambda_j|.$$

If we substitute this expression into (4) and integrate both parts over, $\nu_1, \dots, \nu_{\frac{N(N-1)}{2}-N}$, then we'll get the joint density distribution for eigenvalues of the matrix:

$$P(\lambda_1, \dots, \lambda_N) = \exp\left(-\sum_{j=1}^N \left(a \cdot \lambda_j^2 - b \cdot \lambda_j - c\right)\right) \prod_{1 \le j < k \le N} |\lambda_k - \lambda_j|.$$

Having replaced $\lambda_j = \left(\frac{1}{\sqrt{2a}}\right) x_j + \frac{b}{2a}$, the joint density distribution will be proportional to the expression:

$$exp\left(-\sum_{j=1}^N \frac{x_j^2}{2}\right) \prod_{1 \le j < k \le N} |x_j - x_k|.$$

In this case the density distribution for the distances between neighboring eigenvalues, $s = \lambda_{k+1} - \lambda_k$, of a random symmetric matrix is defined as follows [10]:

$$P(s) = \frac{\pi s}{2} exp\left(-\frac{\pi}{4} s^2\right). \qquad (5)$$

The eigenvalues should be transformed in the way that their distribution approached the uniform one using the procedure of Gaussian expansion [19].

The random $p \times q$ matrix in asymptotic form, $p, q \to \infty$, with constant relation $\frac{p}{q}$, and standard deviation of the matrix elements, $\sigma \to 0$, such that the limit, $\sigma^2 q$, is finite, exhibits the following distribution of eigenvalues [20]:

$$P(\lambda) = \frac{1}{2\pi\lambda\sigma}\sqrt{\left(\lambda_{max}^2 - \lambda^2\right)}\left(\lambda^2 - \lambda_{min}^2\right), \quad (6)$$

for, $\lambda_{min} < \lambda < \lambda_{max}$, and is zero otherwise; here, $\lambda_{max} = \sqrt{2}\sigma\sqrt{\left(\frac{p+q}{2}\right) + \sqrt{pq}}$, $\lambda_{min} = \sqrt{2}\sigma\sqrt{\left(\frac{p+q}{2}\right) - \sqrt{pq}}$. Thus, all eigenvalues of correlation matrix satisfying the conditions given above are positive and restricted in their values. Also for a correlation matrix the following equality holds: $p=q=N$. Exceeding the interval is possible for finite N, though substantial exceeding by eigenvalue the value of da_{max} means deviation from theoretical results produced in suggestion of independence of the random values. That is why such exceeding points out to dependency between observations in the sense of the correlation coefficients for which the matrix was built.

5 Expanded multivariate model

To create a model for multivariate financial system functioning under influence of external factors and dependent on them financial instruments we hired a model on the basis of combined marginal distributions and copulas for the basic observable variables. Further on the model was expanded with dependent variables. As dependent variables the financial derivatives were used. It should be noted that the model is not limited by the independency requirement for the basic factors; their joint distribution can be written in the form:

$$H(x_1, \ldots, x_n) = P[X_1 \le x_1, \ldots, X_n \le x_n] = C(F_1(x_1), \ldots, F_n(x_n)),$$

where, $F_1, \ldots, F_n$ are marginal distribution function for separate risks; C is n-copula, that characterizes the dependency structure between the risks.

Consider as an example Archimedean copula that can presented in the form: $C(u_1, \ldots, u_n) = \varphi^{[-1]}(\varphi(u_1) + \cdots + \varphi(u_n))$. The universal copula generating algorithm is based upon the following representation:

$$C(u_1, \ldots, u_n) = C_n(u_n | u_1, \ldots, u_{n-1}) \ldots C_2(u_2|u_1)C_1(u_1),$$

where, C_2, is a copula for the first k components, $C_1(u_1)=u_1$. To generate the data from the joint distribution the following steps should be performed:

- generate n independent random values, $\nu_1, \ldots, \nu_n$;
- then sequentially compute the following values: $u_1 = \nu_1$, $u_2 = C_2^{-1}(\nu_2|u_1)$, ..., $u_n = C_n^{-1}(\nu_n|u_1, \ldots, u_{n-1})$.

The algorithm presented is simplified for Archimedean copulas, for example, to the following:

$$u_2 = \varphi_2^{[-1]}\left(\varphi_2\left(\varphi_1^{[-1]}\left(\frac{\varphi_1(u_1)}{\nu_2}\right)\right) - \varphi_2(u_1)\right).$$

The model application is oriented to elliptical, Archimedean, and extreme value copulas. The right tails of the marginal distributions are described by the generalized Pareto distribution of the form:

$$GPD_{\xi,\beta}(x) = \begin{cases} 1-\left(1+\frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1-\exp\left(-\frac{x}{\beta}\right), & \xi = 0, \end{cases}$$

where $\beta > 0$, and $x \geq 0$ with $\xi > 0$, and, $0 \leq x \leq \frac{-\beta}{\xi}$, with ξ<0; ξ is parameter of distribution form; β is scale parameter. The marginal distributions of the central observations are described by normal distributions.

6 The derivative financial instruments

To derivative financial instruments are related the contracts, the cost of which depends on some other financial instrument, stock index or actual interest rate. To study the possibility of modeling the financial system with such derivatives as options, forwards, and futures in conditions of large data bases the model should be expanded with actual computed costs according to the methodologies used in practice.

Option is a standard financial document that proves a right to buy (sell) financial instruments (goods, currencies) on predetermined conditions in the future with fixed price related to the time of signing the option agreement or other moment of time according to decision of the contract sides. In practice, for analytical description of currency options the Garmin-Colhagen formula is often used that is a special case of the Black-Scholes expression for determining an option price [21]. The Garmin-Colhagen model is based upon the three basic restrictions:

- absence of tax or restrictions for the market operations;
- invariability of riskless interest rates within the period of a contract;
- actual exchange rates for currencies accept random values having lognormal distribution with constant standard deviation σ.

The third restriction influences joint distribution of costs and can create a source for generating internal influence factors in financial system. The option cost for purchasing (call option) can be is computed as follows:

$$V_C(x, K, \sigma, r_d, T, r_f) = xe^{-r_f T} Normal(d_1) - Ke^{-r_d T} Normal(d_2), \tag{7}$$

and the option cost for selling (put option) is determined via the expression:

$$V_P(x, K, \sigma, r_d, T, r_f) = -xe^{-r_f T} Normal(-d_1) + Ke^{-r_d T} Normal(-d_2),$$

where V_C is theoretical option cost for purchasing; V_P is theoretical option cost for selling; x is current exchange rate; K is exchange rate used for creating an option; T is time to the end of option; r_d is riskless interest rate for the first currency; r_f is riskless interest rate related to the second currency; σ is standard deviation of the exchange rate. *Normal* – means a distribution function for normal distribution; the coefficients, d_1, d_2, are determined as follows:

$$d_1, d_2 = \frac{ln\frac{x}{K} + \left(r_d - r_f \pm \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}.$$

The forward contract is a standard document that proves obligations of a person to purchase (sell) stocks, goods or currency at some predetermined moment of time in the future, and on predetermined conditions with the price fixed at the moment of signing the contract. The cost of a forward contract for a selected currency is calculated according to the following formula [22]:

$$V_t = xe^{-r_f T} - Ke^{-r_d T}, \tag{8}$$

where, x, K, r_f, r_d, T, are the same parameters that were used for calculating an option cost.

The futures contract is a standard document that proves an obligation to purchase (sell) stocks, goods or currency at the predetermined moment of time and on predetermined conditions in the future with fixed price at the moment of fulfilling the obligations by the contract sides. The futures of Eurodollar type are the futures contracts with averaged interest rate according to the interbank credits LIBOR (*London Interbank Offered Rate*). There also exist the futures contracts for other currency pairs. The Euroyen futures are nominated in Japanese yens, the Euroswiss are nominated in Swiss francs etc. These contracts are based upon three-month interest rate and are distinguishing by the terms of their action from several months to tens of years. The contract cost is computed via the following formula:

$$V_t = 10000 \times [100 - 1{,}25 f_t],$$

where, f_t, is interest rate; the coefficient 0,25 is related to the three-month contract. To model the systems with the futures contracts of Eurodollar type the model should take into account the dependency structure for the currency rates, and the LIBOR interest rate as a separate random value. The traditional futures contracts have a cost formula similar to the forward contracts, and the use of the formula provides the model with the same features that are characteristic for the forward contracts only.

7 Computational experiment

As statistical data the following daily exchange rates were taken: US dollar (USD), English pound (GBP), Swiss franc (CHF), and Japanese yen (JYN) with respect to euro (EUR) for the period from 2000 to 2007. The joint exchange rates distribution model was constructed on the basis of Gumbel copula.

To develop experimentally multivariate model the futures and options were used as derivatives with linear and nonlinear costs with respect to the basic financial instruments. In the formulae for cost the following riskless values of interest rates proposed by central banks were used: JYN 0.5%, CHF 2.75%, GBP 5.0% (*Bank of England Bank Rate*), USD 3.25% (*Federal Funds Rate*), EUR 3.0% (*Eurozone Refinancing Rate*).

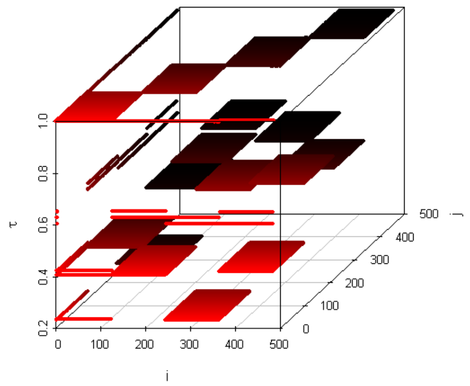

Fig. 1. The matrix of Kendall rank correlation coefficients for the noiseless model

The main point of the study was to consider the dependences that are available in such system of financial instruments; that is why the total contract sums were normalized to the unit of respective currency. The normally distributed zero mean random values were added to the cost of derivative instruments using (7) and (8). The standard deviation of the random value was selected to be equal to 0.1 of the standard deviation for the cost of each instrument. To each exchange rate were added 60 forwards and 60 options with different future costs and final terms of the contracts. For the samples compiled this way with adding noise and without it the empirical 484×484 matrices were computed containing linear correlation coefficients, ρ; linear Kendall rank correlation coefficients, τ (Figs. 1 and 2); and rank correlation Spearman coefficients, ρ_S.

The comparison was performed for the empirical density of the matrix eigenvalues distribution:

$$P(\lambda) = \frac{1}{N}\frac{dn(\lambda)}{d\lambda},$$

where, $n(\lambda)$, is a number of matrix eigenvalues that are less than λ, with theoretical distribution of eigenvalues under suggestion of randomness of matrix (6). For all dependency measures the main bulk of the eigenvalues corresponds to theoretical restrictions. Four of the eigenvalues in all cases exceed the theoretically maximum values for all empirical correlation matrices. For example, the matrix of linear correlations contained, λ_{max}=8.8475. Generally, the four maximum eigenvalues were as follows:318.7, 69.6, 16.6, and 12.5, and all other 480 eigenvalues exhibited positive values less than 1.02.

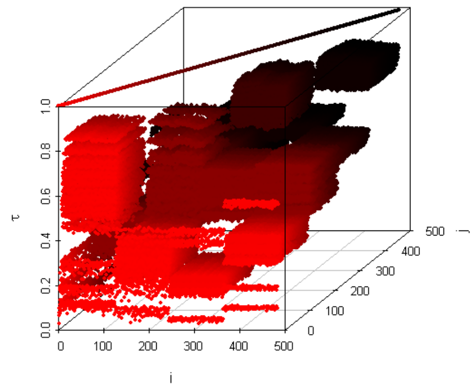

Fig. 2. The matrix of Kendall rank correlation coefficients for the model with noise

For the empirical correlation matrices was computed empirical distribution of distances between the eigenvalues transformed via Gaussian expansion and theoretical distribution of distances for the corresponding symmetric random matrix in (5). The empirical distributions for correlation matrix and theoretical distributions for the random matrices turned out to be the same for the main bulk of eigenvalues except for the largest eigenvalue of empirical correlation matrix. This eigenvalue was much larger in comparison to the theoretical distribution. For the linear correlations it was equal to the level of, 99.9992%. In the right tail of the distribution 5% of the eigenvalues exceed theoretical threshold for 95% of observations, and theoretical threshold of 97% exceed 2.7% of the eigenvalues (Fig. 3).

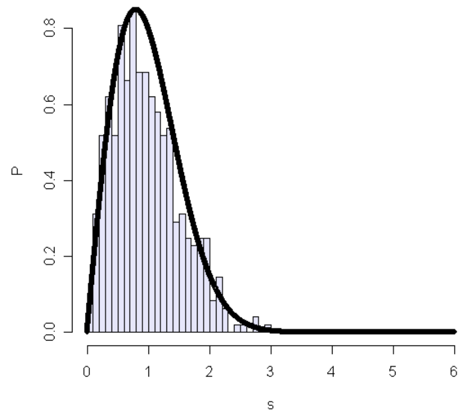

Fig. 3. Empirical density of distribution for distances between the eigenvalues of empirical matrix of linear Pearson coefficients and theoretical density of distribution for distances

As a result of performing the computational experiments it was established the following: the distributions of eigenvalues and distances between the eigenvalues for empirical correlation matrices with the use of different dependence measures demonstrated quite similar behavior. The computed distributions of the eigenvalues provide a possibility for further correct determining the principal factors for constructing adequate models for the processes under consideration and risk estimation.

8 Conclusion

As a result of the studies performed it was established that the spectra of relatively small eigenvalues for the linear correlation coefficients, ρ, rank correlation Kendall coefficients, τ, and rank correlation Spearman coefficients, ρ_S, in a case of adding some noise are practically similar to the theoretical spectra of random matrices. That is why the optimal Markowitz portfolio that corresponds to small eigenvalues should be compiled only after filtering of statistical data.

The number of eigenvalues that exceed theoretical thresholds corresponds to the principal factors in a model. This result provides a possibility for determining correctly the number of principal factors to construct mathematical models for practical applications. The difference between theoretical and empirical distributions of distances between eigenvalues means that in practice there almost always exist a large eigenvalue indicating (in economic interpretation) on existence of dominating generalized market factor.

It was also established that no extra internal influence factors exist when the widely used models of derivative costs are hired. This is a positive sign for carrying out simplified computations.

To our opinion, it would be logically concentrate the future research on refining the results achieved for computing theoretical distributions of the eigenvalues and the distances between them for symmetric positively defined matrices, the elements of which are restricted with the interval of, [-1,1]. The problem to be solved is also touching upon investigation of influence of nonlinear strictly increasing transforms on the distributions of empirical dependency measures matrix eigenvalues. And, it would also be interesting to study the dependency measure in the form of Matsusita-Hellinger metrics as well as compile other possible measures.

References

1. Denuit M., Dhaene J., Goovaerts M., Kaas R.: Actuarial theory for dependent risks. John Wiley & Sons, Ltd, New York (2005)
2. Trofymchuk O.M., Bidyuk P.I., Prosyankina-Zharova T.I., Terentiev O.M.: DSS for modeling, forecasting and risk estimation. LAMBERT Academic Publishing, Kyiv (2019)
3. Rasmussen C.E., Williams C.K.I.: Gaussian processes for machine learning. The MIT Press, Cambridge (2006)
4. Malevergne Y., Sornette D.: Collective origin of the coexistence of apparent RMT noise and factors in large sample correlation matrices. Physica A: Statistical Mechanics and its Applications, vol. 331, 2004, Issues 3-4, pp. 660–668. DOI:10.1016/j.physa.2003.09.004

5. Noh J.: A model for correlations in stock markets // Physical Review E (2000). V.61, pp. 5981 – 5982.
6. Fama E., French K.: Multifactor explanations of asset pricing anomalies. Journal of Finance. (1996). vol.51, no.1, pp. 55 – 84.
7. Mehta M.: Random Matrices. 3ed. Elsevier Academic Press, Amsterdam (2004)
8. Alon N., Krivelevich M., Vu V.: On the concentration of eigenvalues of random symmetric matrices. Israel Journal of Mathematics. V.131, pp. 259 –267. (2002). DOI: 10.1007/BF02785860
9. Laloux L., Cizeau P., Potters M., Bouchaud J.: Random matrix theory and financial correlations. International Journal of Theoretical & Applied Finance. no. 3, pp. 391 – 397. (2000). DOI: 10.1142/S0219024900000255
10. Plerou V., Gopikrishnan P., Rosenow B., Amaral L., Stanley H.: Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series. Phys. Rev. Letters, vol.83, no. 7, pp. 1471 – 1474. (1999). DOI: 10.1103/PhysRevLett.83.1471
11. Sharifi S., Crane M., Shamaie A., Ruskin H. Random matrix theory for portfolio optimization: a stability approach. Physica A, vol. 335, no. 3-4, pp. 629 – 643. (2004). DOI:10.1016/J.PHYSA.2003.12.016
12. Nelsen R.B.: An Introduction to Copulas. Springer, New York (2006)
13. Granger C., Maasoumi E., Racine J.: A Dependence Metric for Possibly Nonlinear Processes. Journal of Time Series Analysis, vol. 25, no. 5, pp. 649 – 669. (2004). DOI:10.1111/j.1467-9892.2004.01866.x
14. Kurowicka D.: Uncertainty Analysis with High Dimensional Dependence Modelling. Wiley, New Jersey, New York (2006)
15. Markowitz H.: Portfolio Selection. Journal of Finance, vol. 7, no. 1 pp. 77-91. (1952). DOI:10.2307/2975974
16. Pafka S., Kondor I.: Estimated correlation matrices and portfolio optimization. Physica A, no. 343, pp. 623 – 634. (2004). DOI: 10.1016/j.physa.2004.05.079
17. Laloux L., Cizeau P., Bouchaud J.-P., Potters M.: Noise Dressing of Financial Correlation Matrices. Phys. Rev. Lett. vol.83, no. 7, pp. 1467–1470. (1999). DOI:10.1103/PhysRevLett.83.1467
18. Sastry S., Deo N., Franz S.: Spectral Statistics of Instantaneous Normal Modes in Liquids and Random Matrices. Phys. Rev. E, vol. 64, pp. 16305–16309. (2001). DOI: 10.1103/PhysRevE.64.016305
19. Bruus H., Angles d'Auriac J.C.: Energy level statistics of the two-dimensional Hubbard model at low filling. Phys. Rev. B, vol. 55, pp. 9142–9159. (1997). DOI:10.1103/PhysRevB.55.9142
20. Sengupta A., Mitra P.: Distributions of singular values for some random matrices. Phys. Rev. E, vol. 60, no.3, pp. 3389 – 3392. (1999). DOI:10.1103/PhysRevE.60.3389
21. Huckins N.W., Rai A.: Market Risk for Foreign Currency Options: Basle's Simplified Model. Financial Management, vol. 28, no.1, pp. 99-109. (1999).
22. Jorion P.: Financial Risk Manager Handbook, 2nd Edition. Wiley, New Jersey (2003)