The Minimal Ellipsoid and Robust Methods in the Optimal Portfolio Problem

Danilova N., Yao K.

Abstract—**In the paper, the optimal portfolio problem is considered. The difficulty of the problem is that the probability measure by which the mean vector and the covariance matrix are calculated is unknown. Instead of the unknown measure, we can use the training sample, based on which it is possible to evaluate the mean vector and the covariance matrix. We use three methods for evaluation. The first is the MCD method. The second is the robust optimization method. In the third method, the Wasserstein distance is used. Computational examples and analyses of the results are given.**

Index Terms—**Optimal portfolio, machine learning, robust method, Wasserstein distance, mean vector, covariance matrix, cluster.**

I. Introduction

THE problem of optimal portfolio finding has been relevant since the publication of the work of Markowitz in 1952 to the present day.

The investor has n assets, that are traded on the market. The price of the i^{th} asset at the moment of time t is calculated by the recursive formula $S_t^i = S_{t-1}^i r_t^i$. We assume that the random vectors $R_t = \begin{pmatrix} r_t^1 \\ \cdots \\ r_t^n \end{pmatrix}$ are independent and equally distributed, just like the random vector $R = \begin{pmatrix} r_1 \\ \cdots \\ r_n \end{pmatrix}$. The vector R will be called the vector of profitabilities. The paper considers a one-step portfolio $y = \begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix}$, in which capital at the initial moment of time t_0 is $X_0 = \sum_{i=1}^n y_i S_0^i$, at the final moment of time t_1 is $X_1 = \sum_{i=1}^n y_i S_1^i = \sum_{i=1}^n y_i r_i S_0^i$. Coefficients y_i– real numbers, equal to the number of units of the i^{th} asset, included in the investment portfolio. Obviously, the final capital of the portfolio $X_1 = \left(\sum_{i=1}^n x_i r_i\right) X_0$, where $x_i = \frac{y_i S_0^i}{X_0}$ and $\sum_{i=1}^n x_i = 1$. Further, the portfolio will be called the vector $x = \begin{pmatrix} x_1 \\ \cdots \\ x_n \end{pmatrix}$, and the scalar product (R, x) is the portfolio profitability. The problem is to choose the optimal vector (portfolio). A detailed analysis of the optimal portfolio problems can be found in the works [2] and [3].

Markowitz diversification. In the interpretation of Markowitz, the quality of the portfolio is determined by two parameters, the average profitability and the risk. The average profitability is calculated as mean $E(R, x) = (\bar{R}, x)$, $\bar{R} = ER$. The risk is variance $D(R, x) = (\bar{C}x, x)$, $\bar{C}$ is covariance matrix, and $\bar{C} = ERR^T - \bar{R}\bar{R}^T$. In addition to the variance, some other functional also can be used as indicators for risk evaluation [4], [5], [6]. The portfolio should be chosen in a way that the average profitability as high as possible, while the risk as low as possible. Based on this circumstance, the optimal portfolio problem belongs to the optimization problem with vector criteria, which solution usually gets across as the Pareto set. There is one way for calculating Pareto optimal portfolios is the vector criteria scalarization:

$$\max_x [(\bar{R}, x) - \lambda\sqrt{(\bar{C}x, x)}] \tag{1}$$

$$\sum_{i=1}^n x_i = 1.$$

The objective function includes the positive parameter λ with a predetermined set of values. The difficulty of the problem is the mean vector and the covariance matrix calculation are unknown in the probability measure P. Instead of the unknown measure, we can use the training sample $V =< R_1, R_2, \cdots, R_N >$ to evaluate the parameters $\bar{R}$ and $\bar{C}$. For example, if the unknown distribution is a normal distribution, the maximum likelihood estimates of $\bar{R}$ and $\bar{C}$ are noted as $\widetilde{R}$ and $\widetilde{C}$, where $\widetilde{R}$ is sample mean and $\widetilde{C}$ is sample covariance matrix. The substitution $\widetilde{R}$ and $\widetilde{C}$ in (1) instead of $\bar{R}$ and $\bar{C}$ makes it possible to find the sample-dependent solution. If we have another sample, we will obtain other values of $\widetilde{R}$ and $\widetilde{C}$. Therefore, if we use $\widetilde{R}$ and $\widetilde{C}$ instead of $\bar{R}$ and $\bar{C}$ in the problem (1), its solution will not be robust against sample change, that can be discovered in [7], [8], [9], [10]. In [11], the robust formulation of the optimal portfolio problem is considered as one of the ways to obtain the robust decision rule. Namely, the set of possible values of sample means $< \widetilde{R}_1, \cdots, \widetilde{R}_K >$ and the corresponding set of sample covariance matrices $< \widetilde{C}_1, \cdots, \widetilde{C}_K >$ are found. To calculate sample means and sample covariance matrices of these vectors, we use unsupervised learning. Therefore, to obtain the values set of sample means and the values set of sample covariance matrices, divide the sample into clusters is proposed. The maximum likelihood algorithm is proposed [12] to split the sample into two clusters, while the dichotomous algorithm is proposed [13] to split the sample into K clusters. In [11], the problem (1) is transformed into the problem

$$\max_x \min_i [(x, \bar{R}_i) - \lambda\sqrt{(\bar{C}_i x, x)}] \tag{2}$$

$$\sum_{i=1}^n x_i = 1.$$

Manuscript received April 09, 2022; revised October 27, 2022.

N. Danilova is an Assistant Professor of the Department of the Operations Research, Mathematics, Mechanics and Computer Sciences Institute of Southern Federal University, Rostov-on-Don, Milchakova st., 8a, 344090, Russia (e-mail: nvdanilova@sfedu.ru).

K. Yao is a PhD student of Mathematics, Mechanics and Computer Sciences Institute, Southern Federal University, Rostov-on-Don, Milchakova st., 8a, 344090, Russia (e-mail: iao@sfedu.ru).

Other applications of robust optimization methods in financial mathematics can be found in the works [14], [15].

II. Problem Setting. Value at risk approach.

This paper proposes to consider the following formulation of the optimal portfolio problem:

$$\min \alpha \tag{3}$$
$$P(-(R,x) \leq \alpha) \geq \beta,$$
$$(I,x) = 1$$

Let us assume that the sample vectors have a normal distribution with expectation vector $\bar{R}$ and covariance matrix $\bar{C}$. From the normal distribution assumption of sample elements, that the random variable $(R,x) \in N((\bar{R},x),(\bar{C}x,x))$. In this case $P((R,x) \geq -\alpha) = 1 - \Phi(\frac{-\alpha-(\bar{R},x)}{\sqrt{(\bar{C}x,x)}})$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$ is the Laplace function. Therefore, $1 - \Phi(\frac{-\alpha-(\bar{R},x)}{\sqrt{(\bar{C}x,x)}}) \geq \beta$, that is $\Phi(\frac{-\alpha-(\bar{R},x)}{\sqrt{(\bar{C}x,x)}}) \leq 1-\beta$. So, $\frac{-\alpha-(\bar{R},x)}{\sqrt{(\bar{C}x,x)}} \leq \Phi^{-1}(1-\beta)$. From this inequality, we can get $\alpha \geq -(\bar{R},x) - \sqrt{(\bar{C}x,x)}\Phi^{-1}(1-\beta)$. Considering equality $-\Phi^{-1}(1-\beta) = \Phi^{-1}(\beta)$, problem (3) can be rewritten as follow:

$$\max_x((\bar{R},x) - \Phi^{-1}(\beta)\sqrt{(\bar{C}x,x)}), \tag{4}$$
$$(I,x) = 1$$

There is no problem in choosing λ with such problem formulation. Namely, $\lambda = \Phi^{-1}(\beta)$. However, the calculation of $\bar{R}$ and $\bar{C}$ is still a question, let us consider three methods to find them.

III. The robust problem. Robust statistics approach.

In the first method, we will apply the well-known algorithm MCD [16], [17], [18]. Let H_1 as the subsample of the sample V with $|H_1| = K$. Then the sample covariance matrix $C(H_1)$ and the sample mean vector $m(H_1)$ are calculated to this subsample. The ordered permutation π is calculated as $\pi(i) < \pi(j) \leftrightarrow C^{-1}(H_1)(R_{\pi(i)} - m(H_1)), R_{\pi(i)} - m(H_1)) \leq (C^{-1}(H_1)(R_{\pi(j)} - m(H_1)), R_{\pi(j)} - m(H_1)))$.

Based on the permutation π, we form the set $H_2 = \{R_{\pi(i)}, i = 1, \cdots, K\}$.

Let us describe the MCD algorithm.

1. Let us choose the initial subset H_1.
2. Calculate the sample mean $m(H_1)$ and the sample covariance matrix $C(H_1)$. Find the ordered permutation π.
3. Choose the subset H_2.
4. If $\Delta C(H_1) > \Delta C(H_2)$, then $H_1 := H_2$, go to step 2, else stop.

We will get the subsample of K elements H_1, the sample mean $m(H_1)$, and the sample covariance matrix $C(H_1)$ for the subsample H_1 while it stops. Let us take $\bar{R}$ as the mean normal law $m(H_1)$. To estimate the covariance matrix, consider the ellipsoid $(C^{-1}(H_1)(R-m(H_1)),(R-m(H_1))) \leq d$, where $d = \max\limits_{R \in H_1}(C^{-1}(H_1)(R-m(H_1)), R-m(H_1)))$. Let C be the covariance matrix of the normal law, that $C = \alpha C(H_1)$. While we calculate y, fact-based on the probability $P((C^{-1}(H_1)(R-m(H_1)), R-m(H_1))) \leq y) \approx \frac{K}{N}$ and the random variable $(C^{-1}(R-m(H_1)),(R-m(H_1)))$ has a chi-square distribution with L degrees of freedom. Thus, y is the solution to the equation $X_L^2(y) = \frac{K}{N}$. Next, we use equality $\frac{1}{d}C^{-1}(H_1) = \frac{1}{y}C^{-1}$, which follows $C = \frac{d}{y}C(H_1)$. Let us take $\frac{d}{y}C(H_1)$ as $\bar{C}$.

It is crutial to choose K, the number of subsample elements in this method. The defining element of the choice is the breakdown point, which is defined as an extreme ratio $\epsilon^* = \inf\left\{\frac{K}{N} : \sup\limits_{V'} ||T(V) - T(V')|| = \infty\right\}$ of the sample V volume N and statistics $T(V)$. The statistic $T(V)$ is a finite-dimensional vector. The norm used in the definition is the Euclidean norm. The sample V' is obtained from the sample V by distorting K arbitrary elements in an arbitrary way. For example, the indicator $\epsilon^* = \frac{1}{N}$ in this sample mean. This indicator of statistical robustness was introduced in the work [20] and is widely used in robust statistics.

MCD method has large sample size and $\epsilon^* = \alpha, (0 < \alpha \leq 0.5)$ subsample size $K = [N(1-\alpha)] + 1$. This important result is presented in [21].

Example. Consider a sample have a two-dimensional normal distribution with a given covariance matrix C_1 and a given mean vector m. The sample is clogged with elements from the two-dimensional distribution, which is obtained as a result of the transformation $y = A\epsilon + m$, where A is the factor of the Cholesky transformation of the covariance matrix C_2, m is the mean vector of the original normal distribution, ϵ is a vector, consisting of two independent uniformly distributed random variables on the interval $[\sqrt{-3}, \sqrt{3}]$.

The results are presented in TABLES I, II, III. In TABLE I the sample is clogged by 10%, in TABLE II the sample is clogged by 15%, and in TABLE III the sample is clogged by 20%.

The MCD method uses the parameter K, which, in our opinion, determines the accuracy of estimating the covariance matrix. In TABLE I, II, III, shows the dependence of the error in estimating the covariance matrix for a clogged sample by this method. In these tables, the sample size $N = 330$. The error is the Frobenius norm of the difference between the original and the resulting covariance matrices. In the last column of the table, the Frobenius norm of the difference between the covariance matrix C_1 and the covariance matrix of the whole sample is presented.

TABLE I
Dependence of Frobenius norm on α.
(the sample is clogged by 10 %)

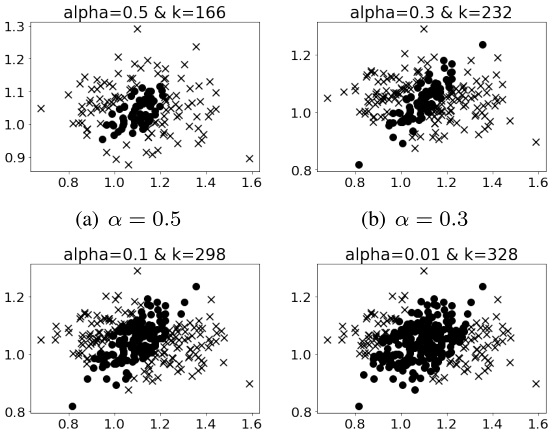
α	0.5 (K=166)	0.3 (K=232)	0.1 (K=298)	0.01 (K=328)	
Frobenius norm	4.4907	1.3692	0.2692	0.0139	0.0619

TABLE II
Dependence of Frobenius norm on α.
(the sample is clogged by 15 %)

α	0.5 (K=166)	0.3 (K=232)	0.1 (K=298)	0.01 (K=328)	
Frobenius norm	5.1012	1.6445	0.4132	0.1232	0.088

TABLE III
DEPENDENCE OF FROBENIUS NORM ON α.
(THE SAMPLE IS CLOGGED BY 20 %)

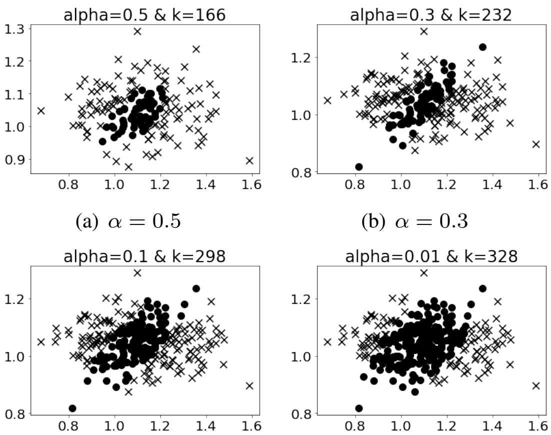
α	0.5 (K=166)	0.3 (K=232)	0.1 (K=298)	0.01 (K=328)	
Frobenius norm	5.729	1.9307	0.5645	0.2399	0.1159



(a) $\alpha = 0.5$ (b) $\alpha = 0.3$

(c) $\alpha = 0.1$ (d) $\alpha = 0.01$

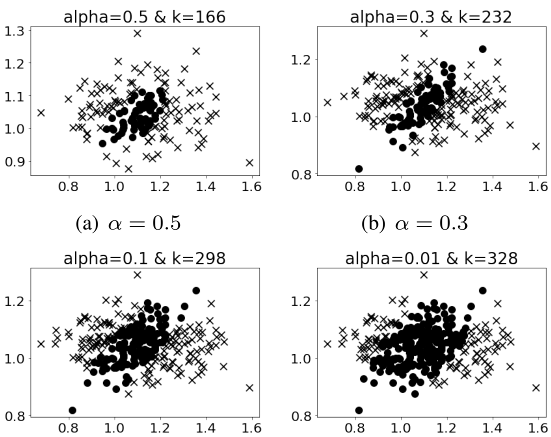
Fig. 1. Illustration of MCD algorithm (the sample is clogged by 10%)



(a) $\alpha = 0.5$ (b) $\alpha = 0.3$

(c) $\alpha = 0.1$ (d) $\alpha = 0.01$

Fig. 2. Illustration of MCD algorithm (the sample is clogged by 15%)



(a) $\alpha = 0.5$ (b) $\alpha = 0.3$

(c) $\alpha = 0.1$ (d) $\alpha = 0.01$

Fig. 3. Illustration of MCD algorithm (the sample is clogged by 20%)

From TABLE I, II, III (columns 2, 3, 4, 5), we may conclude that the Frobenius norm decreases with decreasing α, which indicates the good performance of the MCD algorithm. From column 6 of TABLE I, II, III, we may found that the Frobenius norm of the difference between the covariance matrix C_1 and the covariance matrix of the whole sample is a quite large number, which also looks natural.

Figure 1, Figure 2 and Figure 3 illustrate the operation of the MCD algorithm for a sample, clogged by 10%, 15% and 20%. The elements of the sample that fell into the ellipsoid are dot-shaped. The remaining elements of the sample are cross-shaped.

IV. ROBUST PROBLEM. THE ROBUST PROGRAMMING APPROACH.

Consider the second way to calculate $\bar{C}$ and $\bar{R}$. In the second method, we consider the set of possible values S for $\bar{C}$, the set of possible values T for $\bar{R}$, and the minimax problem statement to obtain the stable portfolio with respect to the sample. That is, we will choose the best portfolio in the worst situation.

The robust problem will look like this:

$$\min_x \max_{\bar{C}\in S, \bar{R}\in T} (\Phi^{-1}(\beta)\sqrt{(\bar{C}x, x)} - (\bar{R}, x)), \tag{5}$$

$$(I, x) = 1$$

An analytical result can be obtained only for simple sets S and T. For example, $T = \{\bar{R} : ||\widetilde{R} - \bar{R}|| \le \epsilon_1\}$ and $S = \{\bar{C} : \bar{C} = \widetilde{C} + \alpha UU^T, 0 \le \alpha \le \epsilon_2^2, ||U|| = 1\}$. Here $\widetilde{R}$ is the sample mean, $\widetilde{C}$ is the sample covariance matrix. With fixed x, it is necessary to find $\max\limits_{\bar{C}\in S}(\bar{C}x, x)$ and $\min\limits_{\bar{R}\in T}(\bar{R}, x)$. Calculating $\max\limits_{\bar{C}\in S}(\bar{C}x, x)$, we use the Cauchy-Bunyakovsky inequality. As a result, we get $(\bar{C}x, x) = (\widetilde{C}x, x) + \alpha(U, x)^2 \le (\widetilde{C}x, x) + \epsilon_2^2||x||^2$. Calculating $\min\limits_{\bar{R}\in T}(\bar{R}, x)$, we also use the Cauchy-Bunyakovsky inequality $(\bar{R}, x) \ge (\widetilde{R}, x) - \epsilon_1||x|$. Thus, the robust optimization problem is as follows:

$$\max_x((\widetilde{R}, x) - \Phi^{-1}(\beta)\sqrt{(\widetilde{C}x, x) + \epsilon_2^2||x||^2} - \epsilon_1||x||) \tag{6}$$

$$(I, x) = 1$$

The objective function is a convex function in this problem. Considering the expression $(\widetilde{C}x, x) + \epsilon_2^2||x||^2$, which we transform as $(\widetilde{C}x, x) + \epsilon_2^2||x||^2 = ((\widetilde{C} + \epsilon_2^2 \,\mathrm{ln})x, x), \mathrm{ln} = \begin{pmatrix} 1 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & 1 \end{pmatrix}$, and coincides with the Tikhonov regularization [22]. The second term of the objective function is related to regularization in machine learning [19].

V. THE ROBUST PROBLEM. THE PROBABILITY APPROACH.

The robust setting of problem (4) in the probability interpretation as follows:

$$\min_x \max_{P\in U_\epsilon(P_0)} (\Phi^{-1}(\beta)\sqrt{(C_P x, x)} - (R_P, x)), \tag{7}$$

$$(I, x) = 1$$

In problem (7), the set $U_\epsilon(P_0) = \{P : D(P, P_0) \le \epsilon\}$, that $D(\cdot, \cdot)$ is the metric on the set of probability measures. Vector $R_P = E_P R$, matrix $C_P = E_P RR^T - R_P R_P^T$. The best known metric between probability distributions P and Q in R^n is the Wasserstein metric, which is defined as $D(P, Q) = \sqrt{\min_\pi E_\pi d^2(X, Y)}, \pi(A, R^n) = P(A), \pi(R^n, B) = Q(B)$.

The Wasserstein metric is used in many works related to find the optimal portfolio, the complete review of these works is given in [23]. Below we restrict the empirical distribution laws, generated by samples of the same size. Usually, the empirical distribution is taken as the initial distribution. When solving the problem (7), we will use the results of the work [24]. We will use the metric $d(X, Y) = ||X - Y||_2$. To simplify problem (7), considering empirical distribution laws $P_X(A) = \frac{1}{n}\sum_{i=1}^n I_A(x_i)$, based on the sample $X = \{x_1, \cdots, x_n\}$, and $P_Y(A) = \frac{1}{n}\sum_{i=1}^n I_A(y_i)$, where $Y = \{y_1, \cdots, y_n\}$ is another sample. The Wasserstein metric $D(P_x, P_y)$ for these distributions is the square root of the solution of the finite dimensional transport problem $\min_\pi \sum_{i=1}^n \sum_{j=1}^n \pi(x_i, x_j) d^2(x_i, x_j)$ with restriction $\sum_{i=1}^n \pi(x_i, y_j) = \sum_{j=1}^n \pi(x_i, y_j) = \frac{1}{n}$. Let us determine $\alpha^2(i) = \min_j d^2(x_i, y_j)$ for each i, and $\bar\pi(x_i, y_j) = \begin{cases} 1/n, j = \alpha(i) \\ 0, j \ne \alpha(i) \end{cases}$ then for each π the inequality $\sum_{i=1}^n \sum_{j=1}^n \pi(x_i, x_j) d^2(x_i, x_j) \ge \sum_{i=1}^n \sum_{j=1}^n \bar\pi(x_i, x_j) d^2(x_i, x_j) = \frac{1}{n}\sum_{i=1}^n \alpha^2(i)$. Therefore, the Wasserstein distance $D(P_X, Q_Y) \ge \frac{1}{n}\sum_{i=1}^n \alpha^2(i)$. By this token the ϵ -neighborhood of the empirical distribution P_0 of the original sample $U_\epsilon(P_0) = \{P_Y : D(P_0, P_Y) \le \epsilon\}$ contained in the set $\bar U_\epsilon(P_0) = \{P_Y : \frac{1}{n}\sum_{i=1}^n \alpha^2(i) \le \epsilon^2\}$. Therefore, the inequality $\max_{P \in U_\epsilon(P_0)} (\Phi^{-1}(\beta)\sqrt{(C_P x, x)} - (R_P, x)) \le \max_{P \in \bar U_\epsilon(P_0)} (\Phi^{-1}(\beta)\sqrt{(C_P x, x)} - (R_P, x))$. Let the elements of the sample $Y = \{y_1, y_2, \cdots, y_n\}$, which generates an empirical distribution law from the set $\bar U_\epsilon(P_0)$, as $y_i = R_i + \alpha_i U_i$, $||U_i||^2 = 1$, $\frac{1}{n}\sum_{i=1}^n \alpha^2(i) \le \epsilon^2$. Let us estimate $\min_{P \in \bar U_\epsilon(P_0)} (x, R_P)$. To do this, consider the equality $(x, R_P) = (x, \widetilde R) + \frac{1}{n}\sum_{i=1}^n \alpha_i (x, U_i)$, further the inequality $(x, \widetilde R) + \frac{1}{n}\sum_{i=1}^n \alpha_i (x, U_i) \ge (x, \widetilde R) - \frac{||x||}{n}\sum_{i=1}^n \alpha_i$. Let us use the concavity of the square root, $\sqrt{\frac{1}{n}\sum_{i=1}^n \alpha_i^2} \ge \frac{1}{n}\alpha_i$. From here, $(x, R_P) \ge (x, \widetilde R) - \epsilon||x||$. Consequently, $\min_{P \in \bar U_\epsilon(P_0)} (x, R_p) \ge (x, \widetilde R) - \epsilon||x||$, and equality is achieved. Let us estimate $\max_{P \in U_\epsilon(P_0)} (C_P x, x)$. It is easy to obtain the following equality for the quadratic form, $(C_P x, x) = (\widetilde C x, x) + \frac{2}{n}\sum_{i=1}^n \alpha_i (R_i - \widetilde R, x)(U_i, x) + (\frac{1}{n}\sum_{i=1}^n \alpha_i^2 (U_i, x)^2 - (\frac{1}{n}\sum_{i=1}^n \alpha_i U_i, x)^2)$. The second term satisfies the chain of inequalities, $\frac{2}{n}\sum_{i=1}^n \alpha_i (R_i - \widetilde R, x)(U_i, x) \le |\frac{2}{n}\sum_{i=1}^n \alpha_i (R_i - \widetilde R, x)(U_i, x)| \le \frac{2}{n}\sum_{i=1}^n |\alpha_i (R_i - \widetilde R, x)(U_i, x)| \le \frac{2||x||^2}{n}\sum_{i=1}^n |\alpha_i| ||R_i - \widetilde R|| \le 2\epsilon||x||^2 \max_i ||R_i - \widetilde R||$. For the last term, the estimate looks like $\frac{1}{n}\sum_{i=1}^n \alpha_i (U_i, x)^2 - (\frac{1}{n}\sum_{i=1}^n \alpha_i U_i, x)^2 \le \epsilon^2 ||x||^2$. Equality is achieved for $U_i = \frac{x}{||x||}$ and for α_i, satisfying the equations $\frac{1}{n}\sum_{i=1}^n \alpha_i^2 = \epsilon^2, \sum_{i=1}^n \alpha_i = 0$. Combining these inequalities, we obtain the estimate $\max_{P \in U_\epsilon(P_0)} (C_P x, x) \le (\widetilde C x, x) + ||x||^2(\epsilon^2 + 2\epsilon \max_{R \in V} ||R - \widetilde R||)$. From problem (7), we obtain the problem as follows:

$$\max_x((x, \widetilde R) - \Phi^{-1}(\beta)\sqrt{(\widetilde C x, x) + ||x||^2(\epsilon^2 + 2\epsilon \max_{R \in V} ||R - \widetilde R||)} - \epsilon||x||), (I, x) = 1 \quad (8)$$

Comparing (6) and (8), we conclude, when using the probabilistic approach, the Tikhonov regularization factor of the covariance matrix exceeds the regularization factor when using robust optimization. The difference depends on the spread of the sample and the mean vector. Comparing all three methods, we found that the MCD method does not require a priori information from the sample. The robust optimization method requires information from the sets of possible parameter values. The latter method requires knowledge of one quantity, which can be estimated from the sample. Therefore, the MCD method and the method, using the Wasserstein distance, have an advantage over the robust optimization method. Based on this, now focus on methods, that have an advantage.

VI. Computational Example

For calculations, the sample of profitabilities ALRS and ROLO for the period from 27.06.2014 to 19.03.2021, volume 298 is used, which is divided into two samples of volume 149 and volume 149, respectively. The first part of the sample (V_1) is used to calculate the portfolio x, and the second part (V_2) is used to calculate the sample mean $\bar V = \frac{1}{|V_2|}\sum_{R \in V_2}(x, R)$. The sample variance $D = \frac{1}{|V_2|}\sum_{R \in V_2}(x, R)^2 - \bar V^2$, and the minimum portfolio profitability $V_{\min} = \min_{R \in V_2}(x, R)$.

Method 1 is the MCD method described above.

Method 2 is the robust method. For the application of this method, we should determine ϵ_1, ϵ_2. To do this, the L pairs of subsamples V_3, V_4 with the same size were randomly obtained from the sample V_1.

For pair V_3, V_4, we calculate mean vectors $m(V_3), m(V_4)$, and calculate the norm of the difference $m(V_1) - m(V_3)$ and the difference $m(V_1) - m(V_4)$. We should choose the maximal value between these two values. Each pair V_3, V_4 corresponds to one value, while L pairs V_3, V_4 correspond to L values. After that, we should choose the maximal value between L values and set ϵ_1 equal to this value.

For each pair V_3, V_4 we calculate covariance matrices $C(V_3), C(V_4)$, and calculate the Frobenius norm of the difference $C(V_1) - C(V_3)$ and the difference $C(V_1) - C(V_4)$. We should choose the maximal value between these two values. Each pair V_3, V_4 corresponds to one value, while L pairs V_3, V_4 corresponds to L values. After that, we should choose the maximal value between L values and set ϵ_2 equal to this value.

Method 3 is another robust method. For the application of this method, we should determine ϵ. The L pairs of subsamples of the same size were randomly obtained from the sample V_1. For each such pair ϵ was calculated by the formula $\epsilon^2 = \frac{1}{L}\sum_{l=1}^L \alpha^2(l)$. Further, the maximum

TABLE IV
DEPENDENCE OF $\bar{V}$, D, V_{min} ON β.

β	0.6	0.7	0.8	0.9
$\bar{V}$(Method 1)	1.000855	1.000880	1.000903	1.000924
$\bar{V}$(Method 2)	0.999016	0.998979	0.998972	0.998935
$\bar{V}$(Method 3)	0.996775	0.996776	0.996777	0.996777
D(Method 1)	0.000977	0.000979	0.000980	0.000981
D(Method 2)	0.001167	0.001168	0.001168	0.001165
D(Method 3)	0.001521	0.001520	0.001520	0.001519
V_{min}(Method 1)	0.864487	0.865256	0.865983	0.866663
V_{min}(Method 2)	0.910368	0.911334	0.911438	0.913030
V_{min}(Method 3)	0.614232	0.614315	0.614390	0.614458

was chosen from the obtained ones. In this example $\epsilon_1 = 0.006717, \epsilon_2 = 0.007110, \epsilon_3 = 0.099896$.

VII. ANALYSIS OF RESULTS

From TABLE IV, we may conclude that the average profitability is better in Method 1, the risk is smaller in Method 1, and the minimal profitability is higher in Method 2.

VIII. CONCLUSION

In this paper, the optimal portfolio problem was considered. Three methods of estimating the covariance matrix and the mean vector were suggested. The first was the MCD method. The second was the robust optimization method. In the third method, the Wasserstein distance was used. Computational examples and analyses of the results also were given.

REFERENCES

[1] Markowitz H. Portfolio selection. J. Financ. vol. 7. n. 1. p. 77–91. 1952.
[2] Brandt M. Portfolio choice problems. Handbook Financ. Economet. n. 1. p. 269–336. 2009.
[3] Steinbach M.C. Markowitz revisited: mean-variance models in financial portfolio analysis. SIAM Rev. vol. 43. n. 1. p. 31–85. 2001.
[4] Konno H., Yamazaki H. Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. Manag. Sci. vol 37. n. 5. p. 519–531. 1991.
[5] Basak S., Shapiro A. Value-at-risk-based risk management: optimal policies and asset prices. Rev. Financ. Stud. vol. 14. n. 2. p. 371–405. 2001.
[6] Rockafellar R.T., Uryasev S. Optimization of conditional value-at-risk. J. Risk. n. 2. p. 21–42. 2000.
[7] Fabozzi F.J., Huang D., Zhou G. Robust portfolios: contributions from operations research and finance. Ann. Oper. Res. Vol. 176. n. 1. p. 191–220. 2010.
[8] Pflug, Pohl. A review on ambiguity in stochastic portfolio optimization. Set-Valued Var. Anal. n. 26. p. 733–757. 2018.
[9] Wozabal D. A framework for optimization under ambiguity. Ann. Oper. Res. vol. 193. n. 1. p. 21–47. 2002.
[10] Pflug G.C., Wozabal D. Ambiguity in portfolio selection. Quant. Finan. vol. 7. n. 4. p. 435–442. 2007.
[11] Belyavsky G.I., Danilova N.V., Logunov A.D. Unsupervised learning and robust optimization in the portfolio problem. Bulletin of higher education institutes. North Caucasus region. Natural sciences. № 4. p.4-9. 2020. (in Russian)
[12] Schlesinger M., Glavach V. Ten lectures on structural and statistical pattern recognition. Naukova dumka. 546 p. 2004. (in Russian)
[13] Rokach Lior, Oded Maimon. Clustering methods. Data mining and knowledge discovery handbook. Springer US. p. 321-352. 2005.
[14] Ben-Tal, Ghaoui, Nemirovski. Robust optimization. Princeton University Press. p. 542. 2009.
[15] Bandi, Bertsimas. Robust option pricing. European Journal of Operational Research. vol. 239. p. 842–853. 2014.
[16] Boudt K., Rousseeuv P., Vanduffel S., Verdonck T. The minimum regularized covariance determinant estimator//Statistics and computing 30,113-128. 2019.
[17] Rousseeuv P., Van Driessen K. A fast algorithm for the minimum Covariance Determinant estimator//Technometrics, 41, 212-223, 1999.
[18] Hubert M., Debruyen M., Rousseeuw J. Minimum covariance determinant and extension//arxiv:1709.07045v1[stat.ME], 2017.
[19] Boyd S., Vanderberghe L. Convex optimization//Campbidge University Press. 716 p. 2009.
[20] Hampel F. A general qualitative definition of robustness//Annul of mathematical statistics. 42, 1887-1896. 1971.
[21] Rousseeum P., Grossmann W., Pflug G., Vincze I. and Wertz W. Multivariate estimation with high breakdown point//Mathematical statistics and Applications, Vol. B, 283-297, 1985.
[22] Tikhonov A., Arsenin V. Methods for solving ill-posed problems. M: Nauka, 430 p. 1974.
[23] Pflug G., Pohl M. A review on ambiguity in stochastic portfolio optimization//Set-Valued and Variational Analysis. v. 26. p. 733-757. 2018.
[24] Blanchety G., Chenz L., Zhoux X. Distributionally robust mean-variance portfolio selection with Wasserstein distances//Management Science. 2021.