B A D A N I A   O P E R A C Y J N E   I   D E C Y Z J E

Nr 1 2008

Grażyna TRZPIOT*, Justyna MAJEWSKA*

INVESTMENT DECISIONS AND PORTFOLIOS CLASSIFICATION BASED ON ROBUST METHODS OF ESTIMATION

In the process of assets selection and their allocation to the investment portfolio the most important factor issue thing is the accurate evaluation of the volatility of the return rate. In order to achieve stable and accurate estimates of parameters for contaminated multivariate normal distributions the robust estimators are required. In this paper we used some of the robust estimators to selection the optimal investment portfolios. The main goal of this paper was the comparative analysis of generated investment portfolios with respect to chosen robust estimation methods**.

Keywords: *Investment decisions, robust estimators, portfolios classification, cluster analysis*

1. Introduction

Nowadays the portfolio analysis is one of the best known and most widely used methods of making investment decisions in capital markets. Following the seminar work by Markowitz (1952) the portfolio selection problem is usually formalized as a mean-risk bicriterial optimization problem where asset expected (mean) return is maximized and some risk measure is minimized.

The classical portfolio theory assumes that the asset return distribution is a multivariate normal. But it is commonly known that leptokurtotic tails of data distribution and contamination of data with outliers are the two features which very often characterize the financial time series. Consequently, in the mean-risk model such parameters as the risk measured usually by standard deviation or variance and the sample mean are quite sensitive to estimation error.

* Katedra Statystyki, Akademia Ekonomiczna, ul. Bogucicka 14, 40-227 Katowice, e-mail: trzpiot@sulu.ae.katowice.pl, majewskaj@wp.pl

** The research has been financed in the year 2007 by the Polish Ministry of Education and Science (research project N111 003 32/0262).

Robust estimators are required to achieve stable and accurate results in case of contaminated data. Numerous of robust estimators are presented and analyzed in literature, so we want to consider some of them in the process of asset selection and their allocation to the investment portfolio.

The main goal of this paper is the classification of generated investment portfolios with respect to chosen robust estimators. Selected methods of cluster analysis were used for the classification. We have tried to isolate homogeneous groups of similar portfolios as well as reveal the relations between these portfolios.

The authors also try to show that applying robust estimation in portfolio analysis ensures better method for effective investment decision-making than the classical portfolio analysis.

The paper is organized as follows. In Section 2, we give a brief overview of the minimum-risk portfolio selection problem. In Section 3 we characterize some robust estimators of risk. In Section 4 we compare and classify portfolios on the basis of various robust estimators.

2. The traditional approach to portfolio optimization

The fundamental goal of the portfolio theory is to optimally allocate investments to different assets. Mean-variance optimization is a quantitative tool, which allows making this allocation by considering the trade-off between risk and return. However, since the covariance matrix can be estimated much more precisely than the expected returns, the minimum variance portfolios (MV) are usually more stable for the composition of the minimum variance portfolio depends only on the covariance matrix of asset returns.

The classical Markowitz optimization problem which constitutes the main theoretical background for the modern portfolio theory is widely described and analyzed in literature, so we will just briefly recall the minimum-variance problem.

For given n risky assets the minimum-variance portfolio (MV) is the portfolio of assets that minimizes the risk measured by the variance of portfolio return for a given covariance matrix C. It is a solution to the following problem:

$$\min_{x=(x_1,\ldots,x_n)^{\mathrm{T}}} x^{\mathrm{T}}Cx \qquad (1)$$
$$s.t. \quad x \in X$$

where $x \in \Re^n$ is the vector of portfolio weights.

The simplest non-empty and bounded set X of feasible portfolios are usually considered as

$$X = \{x \in \Re^n : \sum_{i=1}^{n} x_i = 1, \quad x \geq 0\} .$$

However, the commonly accepted approach to implementation of the mean-risk model is based on usage of specified lower bound μ_0 for expected returns which results in the following minimum risk bounded problem (Ogryczak and Krzemieniowski (2003)):

$$\min\{R(x): \quad \mu(x) \geq \mu_0, \quad x \in X\} \tag{2}$$

where $R(x)$ is the risk measure (risk is measured by volatility of returns).

Various efficient portfolios can be derived by solving problem (2) with changing $\mu_0 \in [\min_{i=1,\ldots,n} \mu_i, \max_{i=1,\ldots,n} \mu_i]$[1]. The efficient frontier is then bounded by the minimum risk portfolio defined as a solution of $\min_{x \in X} R(x)$.

This approach is widely accepted in practice and provides a clear understanding of investor's preferences. Therefore, we also use the bounding approach in our comparison of portfolios.

3. Robust scale estimators

Since the pioneer works of Tukey (1960), Huber (1964) and Hampel (1971) robust statistics are nowadays widely used and new or improved tools are continuously proposed. The aim of robust statistics is to provide tools not only to assess the robustness properties of the classical procedures, but also to produce new estimators and tests that are robust to model deviations.

The breakdown point and the influence function are the most important measures of robustness. But to evaluate the robust estimators the issue of efficiency and of equivariance concepts is very significant (if the estimator is affected by location or scale transformation). So far statisticians have developed various sorts of robust statistical estimators.

Therefore we give below a brief presentation of promising robust estimators of volatility and covariance matrix. We concentrate especially on estimators with i) high breakdown point (near 50%, but for realistic application 20% is satisfactory), ii) property of affine equivariance[2] and iii) fast algorithm to compute them.

[1] For each security $j \in \{1,\ldots,n\}$ its rate of return is represent by a random variable P_j with a given mean $\mu_j = E\{P_j\}$.

[2] Location and scatter estimators $T_n \in \Re^p$ and $C_n \in PDS(p)$ are affine equivariant if and only if $T(AX^n + b) = A \cdot T(X^n) + b$ and $C_n(AX^n + b) = A \cdot C_n(X^n) \cdot A'$ for any vector $b \in \Re^p$ and any non-singular $p \times p$ matrix A.

Some simple scale estimators

The scale of $X = (x_1, ..., x_n)$ [3] is typically estimated by standard deviation

$$\sigma(X) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3}$$

which is very efficient for the assumed normal distribution but highly sensitive to deviation from normality in a sample or empirical distribution. Replacing squares by absolute values and removing the square root leads to the sometimes-used average distance to the average (ADA). But more popular is average distance to the median (ADM)

$$\mathrm{ADM}(X) = \sum_{i=1}^{n}(|x_i - \mathrm{med}(X)|)\ . \tag{4}$$

The most commonly used robust estimator is median absolute deviation about median (the remaining average in (4) is replaced by median) (Hampel, 1974)

$$\mathrm{MAD} = a_n 1.4286 \operatorname*{med}_i \left\{ \left| x_i - \operatorname*{med}_j (x_j) \right| \right\}. \tag{5}$$

The MAD has a simple explicit formula, needs little computation time with 50% breakdown point. However, the MAD has its limitations: low efficiency for data (37%) and an implicit assumption of symmetry.

While the MAD is a location-based estimator as it measures the deviations of the observations from a robust location estimate, the interquartile range is a location-free estimator and is given by

$$IQR = b_n (X_{(3n/4)} - X_{(n/4)})\ , \tag{6}$$

where $X_{(3n/4)}$ and $X_{(n/4)}$ denote the 75th and the 25th percentiles. At symmetric distributions, the IQR has the same influence function as the MAD. But its breakdown point is only 25%. Location-free estimators have the advantage of not implicitly relying on symmetric noise distribution.

Rousseeuw and Croux (1993) proposed two statistics S_n and Q_n as alternatives to the MAD. The first is

$$S_n = c_n 1.1926 \operatorname*{med}_{i=1,...,n} \operatorname*{med}_{j \neq i} |\, x_i - x_j \,|\ . \tag{7}$$

It has significantly higher normal efficiency (58%) and it does not depend on symmetry.

[3] The mean of X is denoted by $\bar{x}$.

The second robust scale estimator Q_n has significantly higher normal efficiency (82%) and it also does not depend on symmetry

$$Q_n = d_n 2.2219\{| x_i - x_j |; i \le j\}_{(k)}, \tag{8}$$

where $k = \binom{h}{2} \approx \binom{n}{2}/4$ with $h = \left\lfloor \frac{n}{2} \right\rfloor + 1$. That is, we take the k-th order statistic of the $\binom{n}{2}$ interpoint distances.

The constants a_n, b_n, c_n, d_n are correction factors that can be chosen depending on the sample size to achieve unbiasedness. Note that S_n and Q_n estimators do not need any location estimate.

Some of the most popular robust estimators are M-estimators (Huber, 1964). We consider the M-estimator of scale S'_n, which is defined for some chosen $c > 0$ as a solution of

$$\frac{1}{n}\sum_{i=1}^{n} \rho_c\left(\frac{x_i - T_n}{S'_n}\right) = \beta , \tag{9}$$

where $\beta = E_\Phi(\rho_c(u))$ and

ρ_c is an even function, $\rho_c(0) = 0$, non-decreasing on $[0,+\infty)$, differentiable a.e.

T_n is auxiliary location parameter calculated usually as med(X).

The function ρ is known as the loss function and "helps" in reducing the effect of outliers. Different $\rho(x, T_n, S_n)$ yields different M-estimators[4] including the standard maximum likelihood estimators.

In our simulation study we use the logistic function due to its good performance in practice

$$\rho_{\log istic}(x) = \psi^2_{\log istic}\left(\frac{x}{0.3739}\right), \tag{10}$$

where $\psi_{\log istic} = \frac{e^x - 1}{e^x + 1}$ and $\rho(\infty) = 1$.

Since $\psi_{\log istic}$ is continuous and increases strictly for positive arguments the solution S_n to (9) always exists and is unique. M-estimator needs to be calculated itera-

[4] For example MAD is robust M-estimator of scale for $\beta = \frac{1}{2}$, $\rho_c(u) = I(|u| > c)$ and $c = \Phi^{-1}\left(\frac{3}{4}\right) = 0.6745$.

tively, starting from the initial scale estimate $S_n^{(0)} = \mathrm{MAD}(X)$ and the iteration steps are as follows

$$S_n^{(k)} = S_n^{(k-1)} \sqrt{2 \frac{1}{n} \sum_{i=1}^{n} \rho_{\log istic} \left(\frac{x_i - T_n}{S_n^{(k-1)}} \right)}$$

until converegence (Werner, 2003).

Another class of estimators related to *M*-estimators is the *S*-estimators (Rousseeuw and Yohai, 1984) which are particularly attractive. An important distinction between these two types of estimators is that *S*-estimators provide simultaneous estimates of both location and scale, while *M*-estimators produce estimates for only one of these quantities. The *S*-estimate of location and scale is defined as location estimate *T* and positive-definite symmetric matrix *C* that jointly solve

$$\min |C|$$

$$\frac{1}{n} \sum_{i=1}^{n} \rho(\sqrt{(x_i - T)^{\mathrm{T}} C^{-1} (x_i - T)}) = K \tag{11}$$

where *K* is a tuning constant[5]. The above equation is often compactly written as

$$\frac{1}{n} \sum_{i=1}^{n} \rho(d_i) = K$$

where d_i is a robust version of the Mahalanobis distance.

To obtain robust estimates the loss function ρ must satisfy the following conditions: i) ρ is symmetric, has a continuous derivative ψ and $\rho(0) = 0$; ii) there must exist $c > 0$ such that ρ strictly increases on $[0, c]$ and constant on $[c, \infty)$; iii) $\psi'(y)$ and $u(y) = \frac{\psi(y)}{y}$ are bounded and continuous and $u(y)$ is non-increasing in $[0, +\infty)$[6]. If loss function ρ satisfies these conditions the *S*-estimator is asymptotically normal, consistent and has bounded influence function (Lopuhaä, 1989).

[5] $K = 6$ or $K = 9$, see Hoaglin, Mosteller, Tukey [3] and Kafadar [5].

[6] Tukey's beweight function satisfies the above mentioned conditons and it is given by

$$\rho(r) = \begin{cases} \frac{c^2}{6}\left(1 - \left(1 - \left(\frac{r}{c}\right)^2\right)^3\right), & |r| \le c, \\ \frac{c^2}{6}, & \text{otherwise.} \end{cases}$$

According to many other authors in section 4 we have chosen Tukey's beweight function as ρ for the simulation study.

The next robust scale estimator for heavy-tailed symmetric distribution is an *A*-estimator presented by Lax in 1985. For given *n* independent and identically distributed observations $X_1, ..., X_n$ the *A*-estimator of scale with biweight function is given by

$$s_A^2 = \frac{k_A^2}{n-1}\sum_{i=1}^{n}(1-u_i^2)^4 e_i^2 \tag{12}$$

with

$$k_A^{-1} = \frac{1}{n}\sum_{i=1}^{n}(1-u_i^2)(1-5u_i^2)\,; \qquad u_i = \frac{e_i}{cs_0}; \qquad e_i = \begin{cases} X_i - M; (|X_i - M|) \le cs_0 \\ 0; (|X_i - M|) > cs_0 \end{cases}$$

where:

M is an *M*-estimator of location,

s_0 is generally taken to be the median absolute deviation of the X_t from the sample median,

c – positive constant[7] that depends on the choice of auxiliary scale estimate $s_0 > 0$.

The next class of estimators are *t*-estimators of volatility which were recommended by Tchernitser and Rubisov (2005). The *t*-estimator for variance/volatility takes the following form:

$$\hat{\sigma}^2 = \frac{1}{n}\left(\frac{\nu+1}{\nu-2}\right)\sum_{i=1}^{n}X_i^2\left(1+\frac{X_i^2}{(\nu-2)\hat{\sigma}_0^2}\right)^{-1}.$$

T-estimators based on t-distribution require iterative procedures and iteratives always starting with an estimate $\hat{\sigma}_0^2$, the MAD would be a natural choice in this case. For $v = 5$ the *t*-estimator performs best.

4. Simulation results

The present section is devoted to an analysis and classification of generated investment portfolios with respect to chosen robust estimation.

[7] According to John Randal Peter Thomson and Martin Lally [9] the *A*-estimators perform best with $c = 10$ or $c = 11$ than the value $c = 9$ found by Lax.

For our experiment we use market indexes: WIG 20, WIRR and MIDWIG of the Polish Stock Exchange Members. WIG 20 index is based on the value of a portfolio of the 20 largest and most liquid companies from the main market. Investments funds and more than 5 companies from one sector (based on the exchange sector classification) cannot participate in the WIG20 index. The MIDWIG index covers maximum 40 mid-size companies listed on the main market. The MIDWIG index excludes companies from the WIG20 index and investments funds. And the WIRR index is the index for the smallest companies of the main market.

We have generated 1000 weekly return following multivariate normal distribution $N(\mu, \sigma)$, where parameters μ and σ were estimated weekly based on the returns of each index WIG 20, WIRR, MIDWIG. For this purpose we used the Monte Carlo simulation. In order to reduce estimation errors we have chosen a weekly periodicity for the rates of return (Simaan, 1997).

We have considered three types of database: without contamination, next we used 2.5 and 5 percentage contamination level. The following two types of contamination have been studied:

i) Substitutive contamination: random replacement of 2.5% of the asset returns by a specific value. This value was calculated based on the estimator of expected return of the asset return plus three times the standard deviation of the corresponding asset return.

ii) Point mass multiplicative contamination: random multiplication of 5% of the asset returns by a specific value – three times the estimated standard deviation. The contamination occurs for each of the three series at the same data points.

For each dataset we have established the risk estimators described in the previous section (see Table 1).

Table 1. Risk estimators calculated for three datasets

Amount of contamination	indexes	Standard dev.	IQR	*MAD*	*A*-estimator	*t*-estimator	Q_n	S_n	ADM	*S*-estimator	*M*-estimator
0%	**WIRR**	0,0287	0,0284	0,0283	0,0286	0,0372	0,0227	0,0285	0,0304	0,0295	0,0281
	WIG20	0,0389	0,0439	0,0400	0,0391	0,0423	0,0314	0,0314	0,0439	0,0414	0,0400
	MIDWIG	0,0232	0,0223	0,0239	0,0233	0,0248	0,0185	0,0234	0,0248	0,0240	0,0233
2.5%	WIRR	0,0399	0,0296	0,0293	0,0316	0,0385	0,0247	0,0300	0,0296	0,0296	0,0292
	WIG20	0,0490	0,0454	0,0411	0,0428	0,0489	0,0342	0,0410	0,0401	0,0413	0,0409
	MIDWIG	0,0289	0,0227	0,0243	0,0257	0,0261	0,0204	0,0243	0,0240	0,0238	0,0225
5%	WIRR	0,0402	0,0304	0,0299	0,0353	0,0394	0,0276	0,0315	0,0301	0,0323	0,0301
	WIG20	0,0564	0,0467	0,0421	0,0429	0,0462	0,0389	0,0426	0,0433	0,0456	0,0399
	MIDWIG	0,0346	0,0230	0,0246	0,0285	0,0271	0,0229	0,0252	0,0259	0,0258	0,0276

Source: Own calculations.

After analysing the results we observed that MAD and Q_n estimators offer a stable estimate of volatilities. *S*-estimator and ADM also offer quite stable results. Whereas, together with increasing the percentage of contamination, the standard deviation changes substantially.

In the next stage we solved the minimum risk model (1) changing every time the estimators of risk. The tables 2–4 present optimal minimum-risk portfolios.

Table 2. Shares of optimal portfolios for dataset without contamination

risk estimator	WIRR	WIG20	MIDWIG	portfolio risk
standard deviation	32,39%	17,61%	50,00%	1,69%
IQR	35,37%	14,63%	50,00%	1,69%
MAD	34,32%	17,12%	48,57%	1,72%
A-estimator	32,58%	17,48%	49,94%	1,70%
t-estimator	27,99%	22,01%	50,00%	1,93%
Q_n	32,83%	17,17%	50,00%	1,35%
S_n	29,96%	25,16%	44,88%	1,62%
ADM	33,84%	16,16%	50,00%	1,82%
S-estimator	33,17%	16,83%	50,00%	1,76%
M-estimator	33,74%	16,61%	49,65%	1,70%

S o u r c e: Own calculations.

Table 3. Shares of optimal portfolios for dataset with 2.5% of contamination

risk estimator	WIRR	WIG20	MIDWIG	portfolio risk
standard deviation	29,90%	20,10%	50,00%	2,18%
IQR	34,99%	15,01%	50,00%	1,73%
MAD	34,00%	17,47%	48,53%	1,76%
A-estimator	32,89%	18,15%	48,96%	1,87%
t-estimator	30,71%	19,29%	50,00%	2,06%
Q_n	33,67%	17,77%	48,56%	1,48%
S_n	32,84%	17,79%	49,37%	1,77%
ADM	32,78%	18,08%	49,15%	1,75%
S-estimator	32,91%	17,09%	50,00%	1,75%
M-estimator	32,99%	17,01%	50,00%	1,69%

S o u r c e: Own calculations.

Table 4. Shares of optimal portfolios for dataset with 5% of contamination

risk estimator	WIRR	WIG20	MIDWIG	portfolio risk
standard deviation	35,18%	18,77%	46,04%	2,38%
IQR	34,52%	15,48%	50,00%	1,72%
MAD	33,74%	17,82%	48,44%	1,73%
A-estimator	31,32%	22,07%	46,61%	1,97%
t-estimator	28,47%	21,53%	50,00%	2,02%
Q_n	33,96%	17,98%	48,05%	1,60%
S_n	32,30%	18,53%	49,17%	1,78%
ADM	35,43%	18,02%	46,55%	1,79%
S-estimator	32,69%	17,31%	50,00%	1,84%
M-estimator	36,39%	21,63%	41,99%	1,81%

S o u r c e: Own calculations.

From the result analysis of Tables 2–4 some conclusions may be drawn:

- portfolios based on Q_n, MAD, IQR and *S*-estimators have the most stable weights in the presence of contamination of data,
- portfolios based on standard deviation are the most sensitive to the presence of contamination,
- changes in MAD-risk and IQR-risk portfolios weights are practically indistinguishable in the cases with 0%, 2.5% and 5% contamination.

We have also solved problem (2) for each level of required rate of return as an additional insight into the comparison of portfolios. The efficient frontiers are presented in Figure 1.

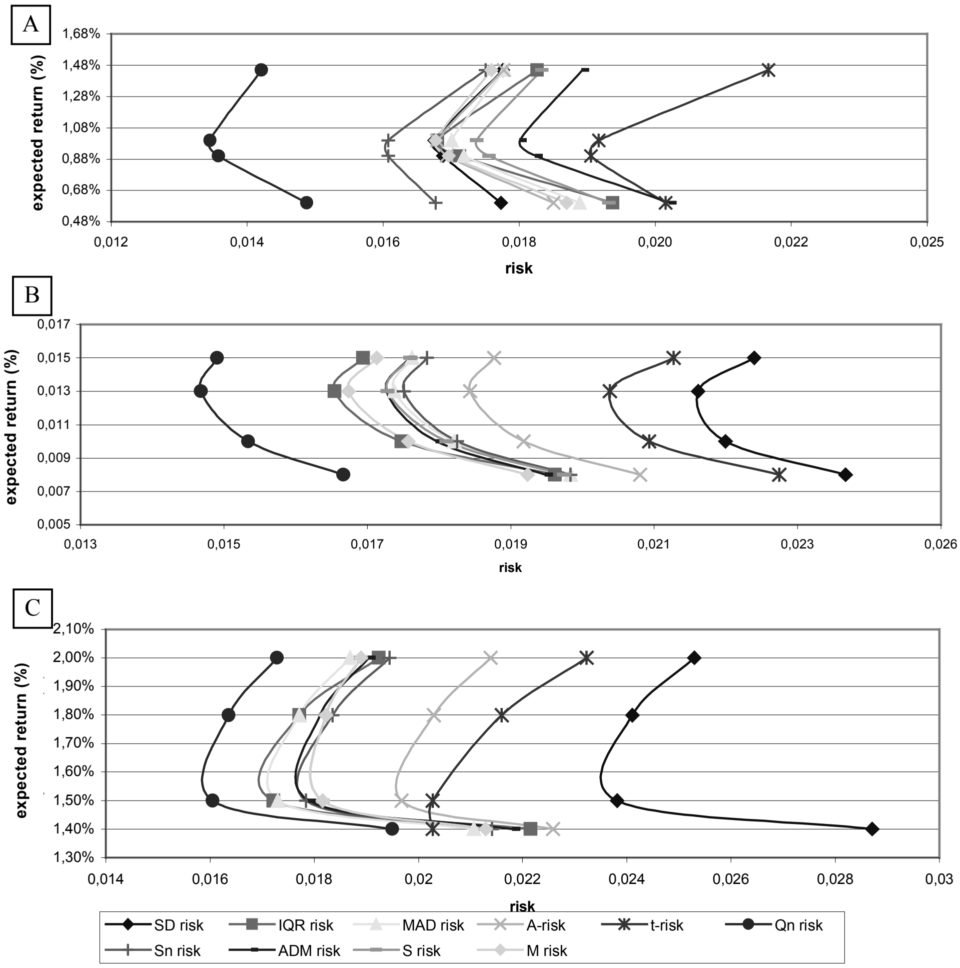

Fig. 1. Efficient frontiers A – for uncontaminated data, B – with 2.5% of contamination and C – with 5% of contamination

We can conclude that model (2) gives the similar results
• with standard deviation, MAD, IQR and *A*-, *M*-, *S*-estimators in the case of Figure A,
• with S_n, MAD, IQR, *M*-, *S*-estimators in the case of the second figure and
• with MAD, IQR, ADM, S_n, *M*-, *S*-estimators in the case of the third figure.

From an investor's point of view it is also interesting to note that portfolio based on Q_n estimator might be classified as the "risk averse" and portfolio based on *t*-estimator as "risk seeking". Also, in the case of 2.5% and 5% of contamination the most risky portfolio is the SD-risk portfolio.

Finally, we have tried to classify the generated investment portfolios with respect to chosen robust estimators. The results are shown as tree diagrams.

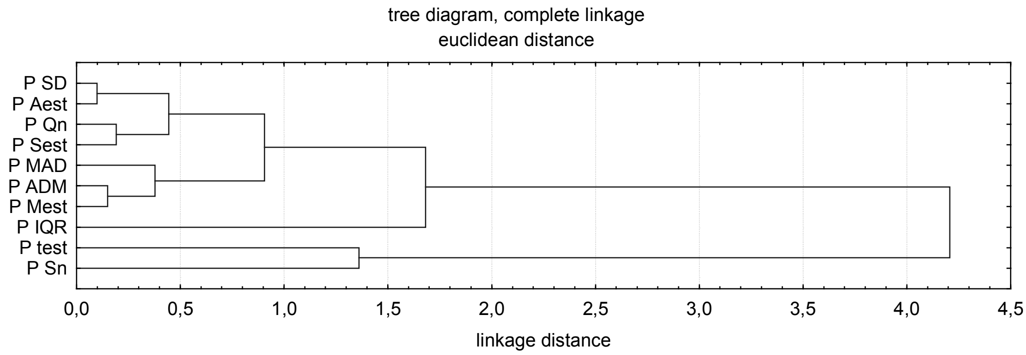

Fig. 2. Tree diagram portfolios based on robust estimators in the case of uncontaminated data

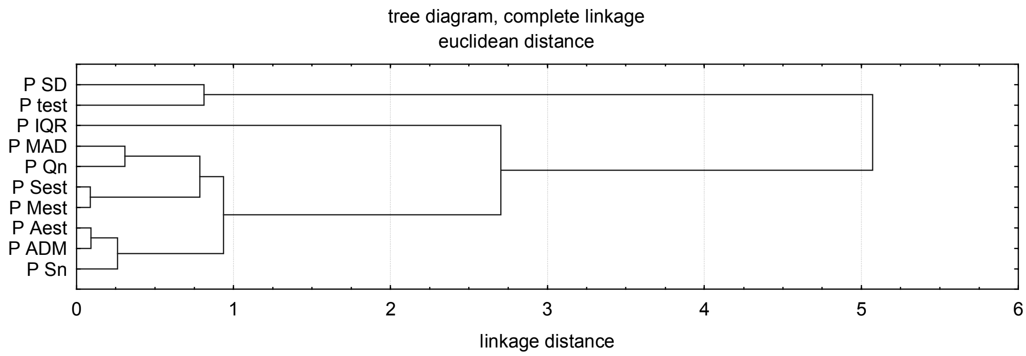

Fig. 3. Tree diagram portfolios based on robust estimators in the case of 2.5% contamination

One may notice homogeneous groups of similar portfolios. Figure 1 presents 4 clusters. Portfolios based on standard deviation, Q_n and *A*-and *S*-estimator. MAD, ADM, *M*-estimator belong to the second group. S_n and *t*-estimator form the third group. And only IQR-risk portfolio forms the last group. Figure 2 presents 4 clusters of portfolios: group 1 – portfolios based on standard deviation and *t*-estimator, group 2 – portfolios based *A*-estimator, ADM and S_n, group 3 – portfolios based on MAD-estimator, Q_n, *S* and *M*-estimator, group 4 – portfolios based on IQR estimator. In Figure 3 there are also 4 clusters: group 1 – portfolios based on standard deviation, *M*-estimator, group 2 – portfolios based on IQR, Q_n, ADM and MAD-estimator, group 3 – portfolios based on S_n, *S*-estimator, group 4 – portfolios based on –-estimator and *t*-estimator.

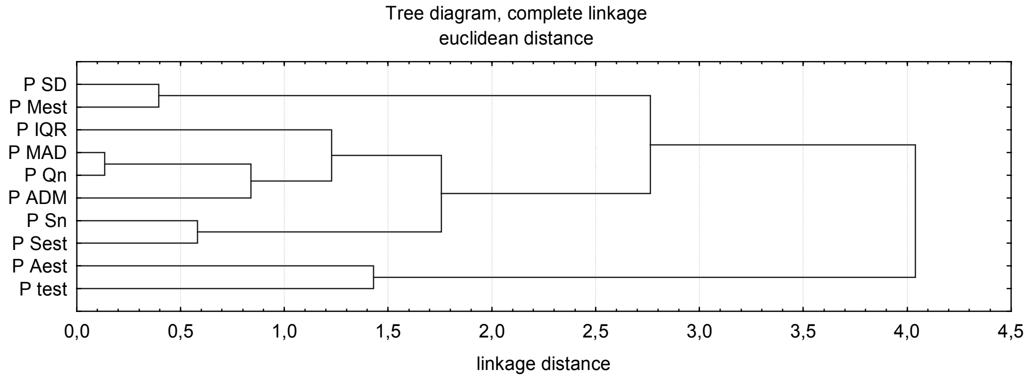

Fig. 4. Tree diagram portfolios based on robust estimators in the case of 5% contamination

5. Conclusions

Robust estimators are the powerful tools for stable evaluation of statistical parameters. In the process of assets selection and their allocation to the investment portfolio the most important is the accurate evaluation of the volatility of the return rate, covariance matrix or correlations.

If volatility is a measure of risk, then MAD and Q_n but also *S*- and ADM estimators are seemed to be the most effective among analyzed volatility estimators. In this paper the homogenous groups of similar portfolios have been obtained as the result of classification. Portfolios based on Q_n and MAD estimators of risk have the most stable weights in the presence of contaminated data, the portfolio based on Q_n estimator might be classified as the “risk averse” and the portfolio based on *t*-estimator as “risk seeking”. Also, in the case of 2.5% and 5% of contamination the most risky portfolio is SD-risk portfolio.

The Achieved results can be used in the investment decisions-making process. These promising results show a need for comprehensive experimental studies analyzing practical performances of the enhanced risk measures.

References

[1] HAMPEL F.R., *A general qualitative definition of robustness*, Annals of Mathematical Statistics, 1971, Vol. 42, No. 6.
[2] HAMPEL F.R., *The influence curve and its role in robust estimators*, JASA 69, 1974.
[3] HOAGLIN D.C., MOSTELLER F., TUKEY J.W., *Understanding robust and exploratory data analysis*, John Wiley, 1983.
[4] HUBER P., *Robust estimation of a location parameter*, Annals of Mathematical Statistics, 1964, 53.
[5] KAFADAR K., *The efficiency of the biweight as a robust estimator of location*, Journal of Research of the National Bureau of Standards, 1983.
[6] LAX D.A., *Robust estimators of scale: finite-sample performance in long-tailed symmetric distribution*, Journal of the American Statistical Association, 1985, 80.
[7] LOPUHAÄ H.P., *On the relation between S-estimators and M-estimators of multivariate location and covariance*, The Annals of Statistics, 1989, 17.
[8] OGRYCZAK W., KRZEMIENIOWSKI A., *On Extending the LP Computable Risk Measures to Account Downside Risk*, Available on the website www.ia.pw.edu.pl
[9] RANDAL J., THOMSON P., LALLY M., *Non-parametric estimation of historical volatility*, Quantitative Finance, 2004, 4.
[10] ROUSSEEUW P.J., CROUX C., *Alternatives to the median absolute deviation*, Journal of the American Statistical Association, 1993, 88.
[11] ROUSSEEUW P., YOHAI V.J., *Robust regression by means of S-estimators. Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics, 26, Springer, New York, 1984.
[12] SIMAAN Y., *Estimation risk in portfolio selection: the mean variance model and the mean-absolute deviation model*, Management Science, 1997, 43.
[13] TCHERNITSER A., RUBISOV D.H., *Robust estimation of historical volatility and correlation in risk management*, University of Toronto, available on website www.gloriamundi.org
[14] TUKEY J.W., *A survey of sampling from contaminated distributions* [in:] I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow, H.B. Mann (eds.), *Contributions to Probability and Statistics*, 1960.
[15] WERNER M., *Identification of multivariate outliers in large data sets* (PhD Thesis), University of Colorado at Denver, 2003.

Decyzje inwestycyjne i klasyfikacja portfeli inwestycyjnych w oparciu o odporne metody estymacji

Zbiory danych finansowych bardzo często charakteryzują się występowaniem wartości wyraźnie różniących się od pozostałych tzw. obserwacji odstających, natomiast rozkład analizowanych danych często jest rozkładem leptokurtycznym z grubymi ogonami. Brak normalności rozkładu oraz występowanie obserwacji odstających w konsekwencji powoduje, że szacowania przy użyciu klasycznych estymatorów są nieefektywne. W tym przypadku zastosowanie znajdują estymatory odporne.

W procesie selekcji aktywów i ich alokacji do portfela inwestycyjnego istotną kwestią jest prawidłowa ocena zmienności stóp zwrotu. Dlatego w niniejszej pracy wykorzystamy szereg odpornych estymatorów zmienności do wyznaczenia optymalnych portfeli inwestycyjnych – portfeli o najmniejszym ryzyku w skład, których wejdą aktywa będące w obrocie na Giełdzie Papierów Wartościowych w Polsce. Głównym celem pracy jest analiza porównawcza oraz klasyfikacja otrzymanych portfeli inwestycyjnych ze względu na wybrane metody estymacji.

Słowa kluczowe: *decyzje inwestycyjne, estymatory odporne, klasyfikacja portfeli, analiza skupień*